UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LETRAS MODERNAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS DA TRADUÇÃO

ALINE MILANI ROMEIRO PEREIRA

TRADUZINDO UM PAÍS:
um estudo da tradução de marcadores culturais a partir de corpora paralelos de
textos português/inglês sobre o Censo Demográfico brasileiro
(Versão corrigida)

São Paulo
2018

ALINE MILANI ROMEIRO PEREIRA

TRADUZINDO UM PAÍS:

um estudo da tradução de marcadores culturais a partir de corpora paralelos de textos português/inglês sobre o Censo Demográfico brasileiro

(Versão corrigida)

Dissertação apresentada ao Programa de Pós-graduação em Estudos da Tradução da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, para a obtenção do título de Mestre em Estudos da Tradução.

Orientadora: Prof^a. Dr^a. Diva Cardoso de Camargo

São Paulo
2018

Autorizo a divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

P436t Pereira, Aline Milani Romeiro
 TRADUZINDO UM PAÍS: um estudo da tradução de
 marcadores culturais a partir de corpora paralelos
 de textos português/inglês sobre o Censo Demográfico
 brasileiro / Aline Milani Romeiro Pereira ;
 orientadora Diva Cardoso de Camargo. - São Paulo,
 2018.
 145 f.

 Dissertação (Mestrado)- Faculdade de Filosofia,
 Letras e Ciências Humanas da Universidade de São
 Paulo. Departamento de Letras Modernas. Área de
 concentração: Estudos da Tradução.

 1. Tradução. 2. Estudos da Tradução Baseados em
 Corpus. 3. Censo demográfico. 4. Tradução de
 marcadores culturais. I. Camargo, Diva Cardoso de ,
 orient. II. Título.

Termo de Ciência e Concordância do (a) orientador (a)

Nome do (a) aluno (a): Aline Milani Romeiro Pereira

Data da defesa: 03/12/2018

Nome do Prof. (a) orientador (a): Diva Cardoso de Camargo

Nos termos da legislação vigente, declaro **ESTAR CIENTE** do conteúdo deste **EXEMPLAR CORRIGIDO** elaborado em atenção às sugestões dos membros da comissão Julgadora na sessão de defesa do trabalho, manifestando-me **plenamente favorável** ao seu encaminhamento e publicação no **Portal Digital de Teses da USP**.

São Paulo, 28/02/2019.

(Assinatura do (a) orientador (a))

PEREIRA, Aline Milani Romeiro. **Traduzindo um país: um estudo da tradução de marcadores culturais a partir de corpora paralelos de textos português/inglês sobre o Censo Demográfico brasileiro.** Dissertação apresentada à Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para obtenção do título de Mestre em Estudos da Tradução.

Aprovado em:

Banca examinadora

Prof. Dr. _____

Julgamento: _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Julgamento: _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Julgamento: _____

Instituição: _____

Assinatura: _____

AGRADECIMENTOS

A Deus, de tudo o princípio e o fim.

Aos meus pais, Anailde e João, por tantos motivos que seria impossível enumerar aqui.

Aos meus irmãos, Eveline e Igor, pelo suporte emocional, financeiro e acadêmico prestado em diversos momentos.

Ao meu marido, Cláudio Santos, pelos momentos compartilhados durante todo este longo processo.

Ao Instituto Brasileiro de Geografia e Estatística, que tem me proporcionado crescimento profissional e oportunidades diversas há quase 15 anos.

Aos meus colegas de trabalho, principalmente aos companheiros tradutores Gisele Flores e La-Fayette Côrtes Neto, pela compreensão e palavras de ânimo.

Aos irmãos da Fraternidade Santo Antônio, por compreenderem minhas limitações e perdoarem os períodos de ausência.

Aos meus amigos, pela torcida e por todos os momentos de alegria que me deram força para seguir em frente.

Aos colegas do TRADUSP, em especial à Flávia Peres Pregmolatto, e à minha xará Aline Vidal por todo o apoio e ajuda.

À professora Diva Cardoso de Camargo, pela orientação cuidadosa, pela atenção, e pelos inspiradores exemplos que me deu ao longo de todo o mestrado. A todos os professores com quem tive a honra de estar em contato durante o curso, em especial às professoras Angela Maria Tenório Zucchi (TRADUSP), Ieda Maria Alves (DFLP), Mariângela de Araújo (DFLP) e Paula Tavares Pinto (UNESP) por suas contribuições nas etapas de qualificação e defesa.

À USP, em especial ao Departamento de Letras Modernas e a todos que, direta ou indiretamente, contribuíram para a conclusão desta etapa em minha vida.

“Traduzir é conviver.”
(João Guimarães Rosa)

RESUMO

Este trabalho apresenta um estudo da tradução para a língua inglesa de textos dos tipos *release* jornalístico e questionário, relativos ao Censo Demográfico 2010 e disponibilizados no *website* do Instituto Brasileiro de Geografia e Estatística (IBGE), com ênfase nas soluções tradutórias dadas aos vocábulos considerados marcadores culturais, pertencentes aos domínios da cultura ecológica, ideológica, material e social (AUBERT, 1981, 2006). Tomando como arcabouço teórico os Estudos da Tradução Baseados em Corpus (ETBC) (BAKER, 1993, 1995, 1996; CAMARGO, 2005, 2007), fundamentos da Linguística de Corpus (BERBER SARDINHA, 2000, 2004), e fazendo uso do programa *WordSmith Tools*, foram analisadas as soluções adotadas, à luz das modalidades de tradução de Aubert (1998), e identificados traços característicos da linguagem da tradução (BAKER, 1996), a saber: explicitação, normalização e simplificação, no corpus traduzido. Também foram feitas comparações com corpora de mesma temática formados por textos originalmente escritos em inglês e obtidos dos *websites* dos institutos internacionais *US Census Bureau* e *Statistics Canada*. Os resultados apontam os domínios da cultura social e ideológica como os mais e os menos frequentes, respectivamente, nas traduções. Os marcadores de maior chavidade foram ‘indígenas’, nos *releases*, e ‘município’, no questionário. Quanto às modalidades, as mais recorrentes nos textos analisados foram: omissão, tradução literal, modulação e transposição. Dada a reduzida presença das características da tradução no questionário, este aspecto foi analisado a partir de exemplos dos *releases*, apenas, sendo a explicitação a marca menos observada, e a normalização, a característica mais presente nos textos. A comparação com os corpora dos institutos internacionais supracitados evidenciou a existência de três palavras-chave coincidentes nos *releases* – ‘*census*’, ‘*population*’ e ‘*household(s)*’ – e de apenas uma palavra-chave no caso do questionário – ‘*person(s)*’. As opções tradutórias adotadas para os marcadores culturais no texto do IBGE não foram encontradas nos corpora comparáveis.

Palavras-chave: Estudos da Tradução Baseados em Corpus; Linguística de corpus; Censo demográfico; Marcadores culturais; Linguagem da tradução.

ABSTRACT

This is a study of the translation into English of texts about the 2010 Population Census – news *releases* and a questionnaire – available from the IBGE (Brazilian Institute of Geography and Statistics) website, with a focus on the translation of the words considered cultural markers, according to Aubert's classification (1981, 2006) into ecological, ideological, social and material cultural domains. Having Corpus-Based Translation Studies (CTS) (BAKER, 1993, 1995, 1996; CAMARGO, 2005, 2007) and fundamentals of Corpus Linguistics (BERBER SARDINHA, 2000, 2004) as its theoretical framework, and based on results provided by the software WordSmith Tools, this research study presents an analysis of the translations of cultural markers according to the concept of translation modalities (AUBERT, 1998), and investigates the occurrence of features of translation (BAKER, 1996): explicitation, normalization and simplification, in the translated corpus. Furthermore, it provides comparisons made with additional corpora about the same topic, formed by texts originally written in English, available from the websites of two international agencies: Us Census Bureau and Statistics Canada. The results show that the social and ideological cultural domains are, respectively, the most and the least common in the translations. The words '*indígenas*', found in the releases, and '*munícípio*', in the questionnaire, had the highest keyness values among the cultural markers analyzed. As for translation modalities, the most recurrent ones in the texts were: omission, literal translation, modulation and transposition. Considering the reduced number of translation features in the questionnaire, that aspect was analyzed in examples taken from the releases, only. In those texts, explicitation and normalization were, respectively, the least and most commonly observed features. The comparison between the IBGE corpus and the ones from the aforementioned agencies revealed that three key-words were found repeated in the releases from the three institutes – 'census', 'population' and 'household(s)' – versus only one key-word in the case of the questionnaire – 'person(s)'. The words used to translate the cultural markers in the texts produced by the IBGE were not found in the comparable corpora.

Keywords: Corpus-Based Translation Studies; Corpus Linguistics; Population census; Cultural markers; Features of translation.

LISTA DE FIGURAS

Figura 1 – Organograma do IBGE	23
Figura 2 – Página inicial do <i>website</i> do IBGE em 1995	30
Figura 3 – Página inicial do <i>website</i> do IBGE em 1998	31
Figura 4 – Página inicial do <i>website</i> do IBGE em 1998	31
Figura 5 – Página inicial do <i>website</i> do IBGE em 1999	32
Figura 6 – Página inicial do <i>website</i> do IBGE em 2005	32
Figura 7 – Página inicial do <i>website</i> do IBGE em 2007	33
Figura 8 – Página inicial do <i>website</i> do IBGE em 2017	33
Figura 9 – Questionário básico do Censo 2010	37
Figura 10 – Questionário da amostra do Censo 2010	38
Figura 11 – Lista de palavras dos <i>releases</i> em português, sem edição.....	55
Figura 12 – Lista de palavras dos <i>releases</i> em inglês, sem edição	56
Figura 13 – Lista de palavras do questionário em português, sem edição	56
Figura 14 – Lista de palavras do questionário em inglês, sem edição	57
Figura 15 – Lista de concordância para ‘aglomerados’ a partir dos <i>releases</i>	57
Figura 16 – Lista de palavras dos <i>releases</i> em português, editada	58

Figura 17 – Lista de palavras dos <i>releases</i> em inglês, editada	58
Figura 18 – Lista de palavras do questionário em português, editada	59
Figura 19 – Lista de palavras do questionário em inglês, editada	59
Figura 20 – Palavras-chave dos <i>releases</i> em português	60
Figura 21 – Palavras-chave do questionário em português	61
Figura 22 – Palavras-chave dos textos do <i>Statistics Canada</i>	63
Figura 23 – Palavras-chave dos textos do <i>US Census Bureau</i>	64

LISTA DE TABELAS

Tabela 1 – Estatísticas gerais dos corpora de estudo obtidas pela ferramenta <i>WordList</i>	65
Tabela 2 – Diferenças percentuais entre as estatísticas gerais das traduções em relação aos textos originais	65
Tabela 3 – Trinta maiores ocorrências das listas de palavras, sem edição, dos <i>releases</i> em português (TO) e em inglês (TT)	68
Tabela 4 – Trinta maiores ocorrências das listas de palavras, sem edição, dos questionários em português (TO) e em inglês (TT)	69
Tabela 5 – Trinta maiores ocorrências das listas de palavras dos <i>releases</i> em português e inglês, após edição	71
Tabela 6 – Trinta maiores ocorrências das listas de palavras dos questionários em português e inglês, após edição	73
Tabela 7 – Percentual de palavras-chave dos <i>releases</i> em relação à língua geral no TO e no TT	75
Tabela 8 – Percentual de palavras-chave dos questionários em relação à língua geral no TO e no TT	75
Tabela 9 – Marcadores culturais identificados entre as palavras-chave dos <i>releases</i> em português, em ordem decrescente de chavicidade e por domínio cultural	76
Tabela 10 – Marcadores culturais presentes apenas na lista de palavras dos <i>releases</i> em português, em ordem decrescente de frequência	77
Tabela 11 – Marcadores culturais identificados entre as palavras-chave do	

questionário em português, em ordem decrescente de chavicidade e por domínio cultural.....	79
--	----

Tabela 12 – Marcadores culturais presentes apenas na lista de palavras do questionário em português, em ordem decrescente de frequência	79
---	----

Tabela 13 – Distribuição percentual dos marcadores culturais por domínio nos <i>releases</i>	80
--	----

Tabela 14 – Distribuição percentual dos marcadores culturais por domínio no questionário	80
--	----

Tabela 15 – Marcadores culturais presentes nos <i>releases</i> originais, com frequências, e respectivas traduções, com frequências e porcentagens de ocorrência nos <i>releases</i> traduzidos para o inglês	81
---	----

Tabela 16 – Marcadores culturais presentes no questionário original, com frequências, e respectivas traduções, com frequências e porcentagens de ocorrência no questionário traduzido para o inglês	103
---	-----

Tabela 17 – Trinta primeiras palavras-chave dos textos em inglês, publicados pelo <i>US Census Bureau</i> , <i>Statistics Canada</i> e IBGE, em ordem decrescente de chavicidade	114
--	-----

Tabela 18 – Quinze primeiras palavras-chave dos questionários em inglês, publicados pelo <i>US Census Bureau</i> , <i>Statistics Canada</i> e IBGE, em ordem decrescente de chavicidade	116
---	-----

LISTA DE ABREVIATURAS E SIGLAS

BNC	British National Corpus
CBG	Conselho Brasileiro de Geografia
CCE	Comissões Censitárias Estaduais
CDDI	Centro de Documentação e Disseminação de Informações
CNE	Conselho Nacional de Estatística
CMGE	Comissões Municipais de Geografia e Estatística
DE	Diretoria Executiva
DGC	Diretoria de Geociências
DI	Diretoria de Informática
DPE	Diretoria de Pesquisas
EJA	Educação de Jovens e Adultos
ENCE	Escola Nacional de Ciências Estatísticas
ETBC	Estudos da Tradução Baseados em Corpus
FUNAI	Fundação Nacional do Índio
GEON	Gerência de Serviços Online
IBGE	Instituto Brasileiro de Geografia e Estatística

INE	Instituto Nacional de Estatística
ISI	International Statistical Institute
RMSP	Região Metropolitana de São Paulo
SIDRA	Sistema IBGE de Recuperação Automática
TO	Texto original
TT	Texto traduzido
UF	Unidade da Federação
UGI	União Geográfica Internacional
UNSD	United Nations Statistics Division
WST	WordSmith Tools

SUMÁRIO

INTRODUÇÃO.....	15
1 O IBGE	20
1.1 A Instituição	20
1.2 As pesquisas populacionais do IBGE e o Censo Demográfico.....	24
1.2.1 O Censo 2010.....	27
1.3 O papel da tradução na disseminação de informações no IBGE	28
1.4 O <i>website</i> do IBGE e o corpus de estudo	29
2 FUNDAMENTAÇÃO TEÓRICA.....	40
1.5 Os Estudos Descritivos da Tradução e os Estudos da Tradução Baseados em Corpus	40
2.2. Os Estudos da Tradução Baseados em Corpus e a Linguística de Corpus	46
2.3 O texto especializado e os marcadores culturais	48
3 METODOLOGIA.....	53
3.1 Composição dos corpora.....	53
3.2 Procedimentos metodológicos.....	54
3.2.1 Obtenção das listas de palavras	55
3.2.2 Listagem das palavras-chave.....	60
3.2.3 Levantamento e classificação dos marcadores culturais.....	62
3.2.4 Análise das traduções	62
4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	65
4.1 Estatísticas gerais.....	65
4.2 Observações acerca das listas de palavras em português e em inglês	67
4.3 Palavras-chave dos textos originais e traduzidos	75
4.4 Seleção, significado e classificação dos marcadores culturais.....	76
4.4.1 Marcadores culturais presentes nos <i>releases</i>	76

4.4.2 Marcadores culturais presentes no questionário	78
4.5 Análise das traduções dos marcadores culturais	80
4.5.1 Análise das traduções dos marcadores dos <i>releases</i>	80
4.5.2 Análise das traduções dos marcadores do questionário	103
4.6 Análise das características da tradução	107
4.6.1 Ocorrências de explicitação	108
4.6.2 Ocorrências de normalização	109
4.6.3 Ocorrências de simplificação	111
4.7 Comparação com textos originalmente escritos em inglês	113
CONSIDERAÇÕES FINAIS	118
REFERÊNCIAS BIBLIOGRÁFICAS	121
APÊNDICE A – Lista de topônimos encontrados nos <i>releases</i> em português, em ordem decrescente de frequência e classificados por tipo.....	126
APÊNDICE B – Marcadores culturais presentes nas <i>KeyWords</i> e <i>WordLists</i> dos <i>releases</i> e questionário não selecionados para análise	130
APÊNDICE C – Marcadores culturais, opções tradutórias da equipe do IBGE e opções tradutórias alternativas sugeridas pelo consultor	132
APÊNDICE D – Data de publicação e títulos dos <i>releases</i> selecionados para o corpus de estudo	134
APÊNDICE E – Ano de publicação e títulos dos textos do <i>Statistics Canada</i>.....	139
APÊNDICE F – Ano de publicação e títulos dos textos do <i>US Bureau of Statistics</i>..	140

INTRODUÇÃO

O conhecimento de uma língua estrangeira – ou ao menos seu satisfatório uso instrumental – é um meio para facilitar a aquisição de conhecimentos, e possibilita o acesso a um universo de informações que estariam inicialmente fora de alcance para uma significativa parcela da população. Também favorece o contato com culturas distantes e a aproximação entre pessoas de modos de vida diversos. O conhecimento da língua inglesa, especificamente, amplia de forma considerável tais possibilidades comunicativas, uma vez que, por razões variadas, este idioma hoje situa-se entre os mais falados no mundo, inclusive por um grande número de não-nativos, como segunda língua. Além disso, o inglês é empregado em contextos diversos, e mantém-se como ferramenta desejada para a comunicação nos meios profissionais e acadêmicos, em nível global.

Dada a relevância do inglês, pode-se concluir que a disponibilização de conteúdos nesse idioma permite projetar em nível internacional atividades realizadas em determinado ambiente, quer para fins meramente recreativos, quer culturais, ou ainda comerciais ou científicos. Assim sendo, a tradução de materiais diversos para a língua inglesa possibilita maior alcance e visibilidade em terras estrangeiras, mesmo em regiões e para um público que não têm o inglês como língua nativa.

Reconhecer o importante papel de um idioma como o inglês implica reconhecer também a relevância da atividade tradutória relacionada a essa língua, uma vez que o trabalho de tradutores se faz necessário para que determinados grupos periféricos tenham acesso a informações universalmente disponibilizadas somente em inglês. De forma análoga, igualmente relevante será a tradução para esse idioma, adotada de modo a promover uma melhor veiculação de informações produzidas em línguas locais para o público estrangeiro, de maneira geral.

Por essa razão, as empresas e instituições brasileiras que pretendem ter alguma inserção internacional em sua área específica de atuação têm investido na divulgação de informações também em inglês. Com o advento da Internet, tal disseminação é facilitada por meio de traduções de *websites* corporativos e institucionais, em uma tentativa de promover a expansão de negócios ou a disseminação de algum trabalho realizado. No caso de fundações públicas, cujas atividades não têm fins lucrativos, mas relacionam-se à prestação de serviços, tal

procedimento também encontra seu lugar. Embora nossa localização geográfica e inserção na realidade sul-americana possam sugerir ser o espanhol o idioma mais frequente, é o inglês que mais favorece nossa comunicação com o resto do mundo, devido às questões de relevância previamente apresentadas.

Dependendo da instituição considerada, a tradução para a língua inglesa pode apresentar um caráter mais ou menos específico em termos de vocabulário, mas é inegável que a tarefa sempre trará diversos desafios, uma vez que tais textos possivelmente constituem-se não somente de informações pertencentes ao universo da instituição, mas compreendem uma gama de características e elementos muito próprios do universo local. Por essa mesma razão, tal conteúdo pode apresentar características bastante peculiares quando expresso em um idioma estrangeiro.

Como exemplo de temática que trata de elementos locais particulares, podemos citar os estudos e pesquisas populacionais. Apesar dos pontos comuns a diversas culturas, os textos a respeito do assunto não raro abordam itens referentes a uma região e ao modo de vida de seu povo, como, por exemplo, o sistema educacional e habitacional, as relações de trabalho, religiões, entre outros.

No Brasil, especificamente, há uma instituição pública responsável pela divulgação de pesquisas e estudos diversos dessa natureza: o Instituto Brasileiro de Geografia e Estatística (IBGE), fundação da administração pública federal vinculada ao Ministério do Planejamento, Desenvolvimento e Gestão, responsável pelo provimento de dados necessários aos mais diversos segmentos da sociedade civil, e também ao governo nas esferas municipal, estadual e federal.

Por conta de sua proeminência como fornecedor de números oficiais que retratam a realidade do país e servem de parâmetro para a adoção de políticas públicas, a instituição encontra-se hoje em posição de relativo destaque entre seus pares de diversas partes do mundo, sendo reconhecida como produtora de informações confiáveis a respeito do Brasil.

De modo a facilitar a divulgação de sua produção além das fronteiras de nosso país, o IBGE mantém versões em inglês de grande parte do conteúdo de seu sítio na Internet (www.ibge.gov.br). É nesse contexto que podemos considerar que, além de produzir material técnico em diversas áreas, o IBGE apresenta ao mundo um panorama da realidade nacional por meio da tradução de notícias e de informações relativas às pesquisas que realiza junto ao povo brasileiro.

Garantir que esse conteúdo esteja traduzido para o inglês significa aumentar as possibilidades de disseminação fora do Brasil das atividades centrais da instituição, o que também contribui para um melhor acesso por parte de usuários estrangeiros às informações mais relevantes sobre nosso país e sua diversidade.

Assim, pode-se dizer que tanto mais fiel será esse “retrato do país” quanto mais for possível expressar em outro idioma os aspectos particulares da vida de nossa população. Para que esse diálogo possa ser bem-sucedido, é necessário que a equipe de tradução responsável pela tradução desses materiais esteja consciente das peculiaridades que envolvem sua atividade profissional e dos diversos fatores que podem aproximar – ou afastar – as informações de seu público-alvo.

Dados o volume e a variedade da produção do IBGE, a prática da tradução no contexto do instituto compreende uma gama de temas passíveis de serem estudados em uma pesquisa acadêmica, uma vez que este trabalha com assuntos de relevância nacional que ajudam a retratar o Brasil, sua vida social, econômica, e sua atuação geocientífica. Sendo necessário, então, eleger um entre os diversos caminhos possíveis para o presente estudo, adotamos como critério para a escolha o grau de alcance da temática entre o povo brasileiro.

Se as questões mais técnicas a respeito da economia, da geografia ou da estatística podem parecer ao público leigo algo de incompreensível, há, na produção do IBGE, um elemento comum conhecido por praticamente toda a população do Brasil: o Censo Demográfico. De modo particular, o Censo aparece neste cenário como carro-chefe da instituição, a atividade sobre qual a maioria dos cidadãos já ouviu falar, ou seja, é a pesquisa de maior apelo entre o grande público. Mesmo que não se tenha conhecimento acerca dos aspectos técnicos relativos ao Censo Demográfico, essa operação é comum a todos os cantos do Brasil, aos mais pobres, aos ricos, a todos os credos, etc., e, entre as atividades conduzidas pelo IBGE, é a que melhor dialoga com todos os brasileiros, a que nos diz mais respeito apenas por sermos quem somos e vivermos onde vivemos. A tradução dos dados relativos ao Censo Demográfico não somente implica a transmissão de informações, mas dá, ao mesmo tempo, visibilidade a elementos bastante particulares de nossa identidade nacional.

Uma segunda e não menos importante motivação para esta pesquisa é a quase total ausência de trabalhos acadêmicos a respeito das traduções realizadas em ambientes governamentais. Embora haja disponibilidade para o público de material

traduzido por e para empresas estatais, este tipo de conteúdo costuma ser bastante ignorado, e os profissionais que o produzem, praticamente invisíveis.

A escolha do material disponível no site do IBGE também se apresenta particularmente interessante por este se tratar de empresa não especializada em tradução. Com a alocação de funcionários efetivos para a realização das traduções a serem veiculadas no seu *website*, o IBGE mostra valorizar a atividade, e também parece buscar, a médio e longo prazos, uma maior constância – e possivelmente melhor qualidade – nos trabalhos de tradução para a língua inglesa.

Considerando o exposto acima, dada a relevância e o alcance dos Censos, assim como suas características peculiares, escolhemos como objeto de estudo nesta pesquisa algumas traduções para o inglês de textos – *releases* jornalísticos e questionário – sobre o Censo 2010 (a edição mais recente) realizadas pela equipe de tradução da página do IBGE na Internet, de modo a alcançar os seguintes objetivos:

Objetivo geral:

- Realizar um estudo descritivo da tradução para a língua inglesa de textos sobre o Censo Demográfico 2010, produzidos pelo IBGE, com ênfase nas soluções tradutórias dadas aos marcadores culturais (AUBERT, 1981, 2006) encontrados no corpus.

Objetivos específicos:

- Identificar os domínios culturais de maior destaque no texto de partida;
- Analisar as traduções dos marcadores culturais;
- Encontrar características da linguagem de tradução presentes nos textos em língua inglesa;
- Fazer comparações entre o vocabulário encontrado nos textos traduzidos para o inglês e em textos sobre a mesma temática originalmente escritos em inglês;

O arcabouço teórico-metodológico adotado baseia-se nos Estudos da Tradução Baseados em Corpus (BAKER, 1993, 1995, 1996); (CAMARGO, 2005, 2007) e nos fundamentos da Linguística de Corpus (BERBER SARDINHA, 2000, 2004). A classificação dos marcadores culturais entre os domínios ecológico, ideológico, material e social seguirá a proposta de Nida (1945) e sua reformulação por Aubert (1981, 2006).

A fim de estudar os marcadores culturais na tradução dos textos escolhidos para nosso corpus, conforme apontado nos objetivos acima descritos, subdividimos a presente dissertação em quatro capítulos, aos quais se seguirão as considerações finais, as referências bibliográficas e os apêndices.

O primeiro capítulo apresenta o Instituto Brasileiro de Geografia e Estatística, abordando sua história e relevância para o país; as pesquisas populacionais realizadas pelo Instituto, e, em especial, o Censo Demográfico; o papel que a atividade de tradução exerce nesse contexto; e informações sobre o corpus de estudo obtido do *website* da instituição.

No capítulo segundo, detalhamos o arcabouço teórico-metodológico, incluindo os Estudos da Tradução Baseados em Corpus, da Linguística de Corpus e questões relativas à tradução especializada e aos marcadores culturais.

O terceiro capítulo traz informações acerca dos materiais utilizados, do programa adotado para a condução da parte experimental da pesquisa, e apresenta, ainda, a metodologia de trabalho para todas as etapas de obtenção e tratamentos dos dados.

Dedicamos o quarto capítulo à apresentação dos resultados obtidos, a análises diversas para os subcorpora dos *releases* e do questionário do Censo 2010 do IBGE, em português e inglês, e às comparações com os subcorpora sobre o tema Censo Demográfico provenientes de outros dois *websites* institucionais, o *US Census Bureau* e o *Statistics Canada*.

As considerações finais apresentam um resumo das análises realizadas, e retomam os questionamentos iniciais motivadores desta pesquisa, ao mesmo tempo em que sugerem possibilidades outras para investigações derivadas do presente estudo.

Assim, a partir dos critérios e etapas acima descritos, optamos por investigar mais detalhadamente este recorte da tradução que o IBGE apresenta ao mundo.

1 O IBGE

A produção de dados oficiais relaciona-se diretamente à capacidade de um país de conhecer sua realidade bem o suficiente para que seja possível tornar cada vez melhor a vida da população. As informações que alimentam tal produção podem ser obtidas de fontes diversas, e processadas de vários modos, permitindo o mapeamento dos aspectos físicos de uma dada nação, de seu potencial na indústria e no comércio, das possibilidades em termos de finanças e de sua forma de organização social.

Diversas nações dispõem de meios e instrumentos que possibilitam a divulgação de tais dados ao grande público, de modo a subsidiar os governos, empresas e demais organismos na criação de medidas voltadas às necessidades da nação e de seu povo. A geração e administração de dados dessa natureza ficam sob a responsabilidade de agências de estatísticas nacionais, de confiabilidade e eficiências garantidas para que possam espelhar devidamente a realidade local.

Não é diferente com o Brasil. Aqui também dispomos de uma instituição cuja principal finalidade é prover dados a respeito da realidade de nosso país. Popularmente conhecido como IBGE, o Instituto Brasileiro de Geografia e Estatística, tem, ao longo de seus 82 anos de atuação, acompanhado a trajetória de nosso país e contado sua história por meio de informações geográficas e estatísticas.

1.1 A Instituição

A trajetória do IBGE tem início no dia 29 de maio de 1936, com a instalação do Instituto Nacional de Estatística (INE), legalmente instituído dois anos antes, por meio do Decreto 24.609, e da criação do Conselho Nacional de Estatística (CNE). O principal propósito dessa organização era integrar e padronizar a produção de estatísticas dos governos municipais e de entidades privadas, com o intuito de construir um “arcabouço dos sistemas de informação do país”, segundo palavras do então presidente Getúlio Vargas.

No ano seguinte, 1937, é criado, com base no mesmo princípio de cooperação interadministrativa, o Conselho Brasileiro de Geografia (CBG), logo incorporado ao INE e já à época autorizado a compor a União Geográfica Internacional (UGI).

Finalmente, em 1938, o Decreto-Lei nº 218 cria a autarquia Instituto Brasileiro de Geografia e Estatística (IBGE), inicialmente formado por dois órgãos colegiados autônomos: o Conselho Nacional de Geografia, novo nome do antigo CBG, e pelo já mencionado CNE.

No ano de 1967, o Instituto passa a se chamar Fundação Instituto Brasileiro de Geografia e Estatística e torna-se subordinado ao Ministério do Planejamento e Coordenação Geral, hoje Ministério do Planejamento, Orçamento e Gestão.

É importante mencionar também que o IBGE faz parte do segmento educacional brasileiro desde 1953, ano em que foi fundada a então Escola Brasileira de Estatística, atual Escola Nacional de Ciências Estatísticas (ENCE). A ENCE tem como objetivo principal formar profissionais na área de Estatística, além de estimular o aperfeiçoamento no ramo por meio dos cursos de pós-graduação *lato sensu* em Análise Ambiental e Gestão do Território, e de fomentar a pesquisa por meio de seus cursos de mestrado e doutorado em População, Território e Estatísticas Públicas.

O IBGE tem estado presente de diversas formas na vida dos brasileiros ao longo dos seus 82 anos de existência. Muitas foram as fases vividas pelo Instituto, cada uma rica em desafios, conquistas e transformações, chegando até o presente momento de considerável estabilidade e indiscutível atuação em prol da população brasileira, segundo estabelece a própria Missão do Instituto: "Retratar o Brasil com informações necessárias ao conhecimento de sua realidade e ao exercício da cidadania." O bom trabalho executado pelo IBGE como prestador de informações oficiais há mais de oito décadas, sempre prezando pela qualidade e sem ingerência governamental sobre suas atividades técnicas, fez do Instituto uma das instituições mais confiáveis e respeitadas no Brasil.

Como principais funções do IBGE nos dias de hoje, podemos elencar as seguintes:

- produção, análise, coordenação e consolidação de informações estatísticas e geográficas;
- estruturação e implantação de um sistema de informações ambientais;
- coordenação dos sistemas estatístico e geográfico nacionais;
- documentação e disseminação de informações.

O IBGE organiza-se em diversos níveis e está presente em todo o território nacional, tendo sua Presidência e Diretoria Executiva localizadas na cidade do Rio de

Janeiro, onde também estão situados outros importantes núcleos: a Diretoria Executiva (DE), a Diretoria de Geociências (DGC), a Diretoria de Pesquisas (DPE), a Diretoria de Informática (DI), o Centro de Documentação e Disseminação de Informações (CDDI), e a Escola Nacional de Ciências Estatísticas (ENCE). Além disso, de modo a garantir que suas atividades tenham a devida abrangência nacional, o IBGE possui uma rede nacional de pesquisa e disseminação, formada por:

- 27 Unidades Estaduais (26 nas capitais brasileiras e 1 no Distrito Federal);
- 27 Supervisões de Documentação e Disseminação de Informações (26 nas capitais e 1 no Distrito Federal);
- 583 Agências de Coleta de dados nos principais municípios.

O corpo técnico do IBGE é atualmente composto por aproximadamente 5000 servidores lotados em todo o Brasil, e merece destaque a diversidade de carreiras encontradas na instituição. Além de funcionários de formação relacionada diretamente aos campos da geografia, economia e estatística, o Instituto congrega uma heterogeneidade de áreas de conhecimento, entre as quais podemos citar: programação e desenvolvimento de *softwares*, *webdesign*, história, biblioteconomia, jornalismo, cinema e vídeo, letras e tradução.

O atual organograma da instituição é apresentado a seguir.

Figura 1 - Organograma do IBGE

Fonte: IBGE

As atividades de tradução não aparecem discriminadas no organograma. Estão subordinadas à Gerência de Serviços Online, setor diretamente vinculado à Coordenação Geral do Centro de Documentação e Disseminação de Informações.

Além de consolidada posição de destaque em todo o Brasil, o IBGE vem alcançando relevância internacional por meio do estabelecimento de diversos convênios com outros países nas áreas de estatísticas, da prestação de consultoria

na condução de pesquisas populacionais, e da participação em organismos internacionalmente reconhecidos. O IBGE esteve, há até pouco tempo, à frente do Instituto Internacional de Estatística (*International Statistical Institute, ISI*), cuja missão é promover a compreensão, o desenvolvimento e as boas práticas em Estatística em todo o mundo. A presidência do renomado instituto foi ocupada, entre os anos de 2015 e 2017, pelo estatístico Pedro Luís do Nascimento Silva, funcionário de carreira do IBGE. Também altamente relevante tem sido a atuação do IBGE como enviado do Brasil à Comissão de Estatísticas das Nações Unidas (*United Nations Statistics Division, UNSD*). Entre março de 2016 e de 2018, nosso país presidiu a Comissão, sendo representado pelo IBGE na pessoa de Wasmália Bivar, ex-presidente e também funcionária de carreira do Instituto. A escolha do IBGE para esta missão confirma que a instituição goza de grande confiabilidade e é respeitada por sua capacidade técnica, não só internamente, mas também fora do país.

O Instituto Brasileiro de Geografia e Estatística se mostra hoje para o Brasil como uma referência sólida ao longo de seus mais de 80 anos de história, e bastante respeitada a despeito da grave crise de confiança que afeta as instituições brasileiras. Esta é também a imagem que espera transmitir em nível internacional, baseada, acima de tudo, na relevância e na qualidade de sua produção.

1.2 As pesquisas populacionais do IBGE e o Censo Demográfico

Para conhecer uma nação é importante conhecer o seu povo. O mapeamento de atividades diversas da população, de seus hábitos, e, principalmente de suas demandas, é realizado pelo IBGE por meio de pesquisas populacionais de temáticas variadas, com níveis de abrangência geográfica diversos e periodicidades variáveis. Dentre as pesquisas populacionais, podemos citar a Pesquisa Nacional de Saúde¹,

¹ Realizada em parceria com o Ministério da Saúde em 2013. Subdivide-se em quatro volumes: Acesso e utilização dos serviços de saúde, acidentes e violências; Ciclos de vida; Indicadores de saúde e mercado de trabalho; e Percepção do estado de saúde, estilo de vida e doenças crônicas.

as Tábuas de Mortalidade² e a Pesquisa Nacional por Amostra de Domicílios³, que fornecem informações a respeito da vida dos brasileiros de diferentes idades e perfis socioeconômicos.

Pesquisas como as mencionadas acima, de recorte temático e metodológico específicos, são muito eficientes para proporcionar uma visão qualitativa do Brasil. No entanto, além de compreender como somos, é necessário saber quantos somos, uma vez que os aspectos demográficos, ou seja, relativos aos tamanhos das populações, também servem para subsidiar a tomada de decisões administrativas e outras:

Uma preocupação fundamental no estudo das populações humanas é com o seu tamanho em determinado momento e com os possíveis fenômenos que determinam ou afetam esse tamanho, tais como os nascimentos, os óbitos e fenômenos migratórios. É importante investigar de que modo cada um desses componentes pode ser afetado por mudanças nos demais e como esses fenômenos se relacionam entre si. (CERQUEIRA e GIVISIEZ, 2004, p.15)

Para realizar tal contagem, e, simultaneamente, levantar informações pertinentes acerca da vida do povo brasileiro, o IBGE realiza, a cada dez anos, a megaoperação pela qual é mais conhecido em todo o país: o Censo Demográfico.

A história dos Censos Demográficos no Brasil antecede a própria criação do IBGE, uma vez que a primeira operação dessa natureza de que se tem registro ocorreu em 1872, e foi chamada Recenseamento da População do Império do Brasil. Diferentemente do que se observa na enorme estrutura que dá suporte à realização do Censo nos dias atuais, as operações censitárias iniciais eram realizadas de forma descentralizada e sem critérios uniformes.

A fase censitária moderna teve início após a criação do IBGE, mais especificamente com o grande recenseamento de 1940, o primeiro realizado pelo Instituto, e que passa a ter sua abrangência temática ampliada a partir da aplicação de um questionário que inclui quesitos acerca de temas como migração interna, emprego e fecundidade, entre outros. Sobre essa primeira edição do Censo decenal,

² Divulgadas anualmente até o dia 1º de dezembro de cada ano, subsidiam o cálculo do fator previdenciário para fins das aposentadorias das pessoas regidas pelo Regime Geral da Previdência Social.

³ Investiga anualmente, de forma permanente, características gerais da população, de educação, trabalho, rendimento e habitação e outras, com periodicidade variável, de acordo com as necessidades de informação para o País.

uma curiosidade: tamanhas eram as dimensões da operação, e com tão grande nível de detalhamento para a cultura da época, que o Censo virou até mesmo tema musical. A letra de um conhecido samba dos anos 40, composto por Assis Valente e interpretado pela cantora Carmem Miranda, então recém-chegada de uma temporada nos Estados Unidos, diz:

Em 1940
lá no morro começaram o recenseamento
E o agente recenseador
esmiuçou a minha vida
que foi um horror
E quando viu a minha mão sem aliança
encarou para a criança
que no chão dormia
E perguntou se meu moreno era decente
E se era do batente ou era da folia

Obediente eu sou a tudo que é da lei
fiquei logo sossegada e falei então:
O meu moreno é brasileiro, é fuzileiro,
e é quem sai com a bandeira do seu batalhão!
A nossa casa não tem nada de grandeza
nós vivemos na pobreza, sem dever tostão
Tem um pandeiro, uma cuíca, um tamborim
um reco-reco, um cavaquinho e um violão

Fiquei pensando e comecei a descrever
tudo, tudo de valor
que meu Brasil me deu
Um céu azul, um Pão de Açúcar sem farelo
um pano verde-amarelo
Tudo isso é meu!
Tem feriado que pra mim vale fortuna
a Retirada de Laguna vale um cabedal!
Tem Pernambuco, tem São Paulo, tem Bahia
um conjunto de harmonia que não tem rival.

Como podemos ver nas palavras do compositor, é objetivo do Censo, de certa maneira, “esmiuçar” a vida e os costumes dos cidadãos, de forma a gerar informações que, tratadas em conjunto, possibilitem conhecer o país e detectar suas demandas, pontos positivos e fragilidades.

Em linguagem simples, o Censo pode ser definido como uma grande operação decenal que conta a população e apresenta informações sobre as condições de vida em cada canto do Brasil. Como carro-chefe do IBGE, o Censo Demográfico também é um dos conteúdos mais acessados do *website* do Instituto, o que confirma seu valor como “retrato do Brasil”:

Em sua já razoavelmente longa história, o IBGE elaborou e produziu um conjunto de pesquisas bastante significativo no que diz respeito à oferta de informações sobre as diversas características demográficas e socioeconômicas da população brasileira. Dentre essas pesquisas, o censo demográfico é um eixo de referência para todas as demais. (OLIVEIRA e SIMÕES, 2005, p. 291)

Os Censos têm sido realizados pelo IBGE a cada dez anos, desde 1940, com exceção do Censo de 1990, que foi adiado em um ano devido à conjuntura político-financeira do país. Logicamente, para que se possa realizar tarefa tão ambiciosa, o processo de preparação dos Censos tem início alguns anos antes da operação propriamente dita. As providências que antecedem o período da coleta de dados são essenciais, uma vez que a boa execução das etapas prévias aponta para o sucesso na condução do Censo. Entre tais etapas temos: atualização dos setores censitários previamente definidos, contratação e treinamento de recenseadores, preparação de campanhas de divulgação e de conscientização a respeito da necessidade de adesão popular ao processo.

1.2.1 O Censo 2010

O Censo de 2010 foi a edição mais recente, na qual mais de 190 mil recenseadores foram às ruas para visitar 67.6 milhões de domicílios os então 5.565 municípios (hoje, 5.570 ao todo) do Brasil. A fase preparatória da operação censitária começou em 2007, e a coleta de dados, com duração de três meses, teve início em 1º de agosto de 2010. A edição de 2010 apresentou avanços em relação aos anos anteriores, sendo os mais significativos: a utilização de uma base territorial única, digital, agrupando as áreas urbana e rural e o Cadastro de Endereços; a utilização do computador de mão equipado com GPS, que facilitou a coleta de dados em áreas rurais e/ou habitadas por populações indígenas. Também é importante destacar a investigação, a partir da edição de 2010, da filiação étnica e da língua falada por tais povos, uma vez que a edição anterior do Censo, de 2000 trazia apenas uma pergunta acerca do pertencimento a tais populações. A inclusão desses quesitos no questionário do censo permitiu, a partir de 2010, abordar de maneira mais particular a sociodiversidade do grupo étnico indígena em nosso país.

1.3 O papel da tradução na disseminação de informações no IBGE

É de se esperar que uma Instituição do porte e importância do IBGE invista fortemente na divulgação de seus produtos⁴, uma vez que, segundo sua missão institucional, um dos objetivos da instituição é o pleno “exercício da cidadania”. Assim, a disseminação de informações assume um papel especialmente relevante para que a sociedade possa tomar conhecimento das informações produzidas a partir dos dados gerados por ela mesma. Por meio de diversas pesquisas do IBGE, as domiciliares, em especial, é possível conhecer a realidade na qual estamos inseridos, à qual influenciamos e pela qual somos afetados. Nesse sentido, podemos dizer que há uma espécie de retroalimentação entre o Instituto e o público que utiliza suas informações, seja em nível acadêmico, governamental ou social.

A unidade responsável por trazer a público os resultados das diversas atividades conduzidas pelo IBGE é o Centro de Documentação e Disseminação de Informações (CDDI), localizado no bairro do Maracanã, na cidade do Rio de Janeiro. O CDDI reúne setores como biblioteca, memória institucional, marketing, produção de vídeos, atendimento ao usuário, loja virtual e o departamento de Internet, a Gerência de Serviços Online (GEON), da qual faz parte o setor de Tradução.

A prestação de serviços de tradução na maior parte das instituições governamentais brasileiras se dá por meio de terceirização, ou seja, da contratação de empresas especializadas ou de profissionais autônomos para responderem pela tarefa. Essa distribuição mostra que a tradução ainda é vista dentro das estatais como atividade periférica ou não integrante do produto final dessas instituições, uma vez que muitos tradutores de fato não fazem parte de fato do “time da casa”.

Podemos dizer que o IBGE está relativamente mais avançado nesse particular. Em 2005 foi realizado um primeiro concurso para a contratação de dois tradutores, para os idiomas inglês e espanhol, respectivamente. Em 2010, um segundo concurso resultou na contratação de mais cinco servidores formados em Letras, para realizar, em princípio, atividades relacionadas à tradução. Apesar desse aparente avanço, o cenário atual está um pouco modificado: por conta da realocação de pessoal e da

⁴ Nomenclatura comumente utilizada pelo Instituto para se referir a todo o material produzido: pesquisas, campanhas, materiais de referência e pesquisa, entre outros.

redistribuição de tarefas, nem todos os servidores inicialmente concursados como tradutores permanecem atuando neste segmento dentro do Instituto.

No nosso caso específico, trataremos das traduções encontradas no *website* do IBGE. A equipe responsável por esse trabalho tem tido atuação constante desde 2005, a despeito da ocasional troca de servidores e da relativa rotatividade de estagiários, uma vez que os estudantes não podem exceder o prazo de dois anos de permanência no serviço federal.

Ter uma equipe própria e fixa para atividade de tradução traz vantagens ao Instituto, dentre as quais destacamos: economia e celeridade nos processos por não precisar buscar periodicamente empresas no mercado; maior comprometimento dos profissionais para com as tarefas realizadas no dia-a-dia; conhecimento mais amplo dos tradutores sobre a dinâmica e a produção da Instituição; menos quebra de continuidade nos processos e projetos. Também por essas razões, e dada a escassez de profissionais qualificados para desenvolver atividades dessa natureza no Instituto, o setor acaba por receber demandas de outras gerências, e atua em projetos além do *website*: tradução de publicações, cartas e documentos diversos, auxílio a outros servidores na elaboração de material para apresentações, tradução para legendagem de vídeos, transcrição de áudios e, ocasionalmente, interpretação consecutiva em reuniões e visitas de estrangeiros.

No caso dos textos relativos ao Censo Demográfico, especificamente, a atividade dos tradutores tem um papel bastante significativo. Disponibilizar os textos sobre o Censo em língua inglesa significa também exhibir o retrato do nosso Brasil. É dizer ao mundo, ou pelo menos a uma significativa parte dele, quem somos e como vivemos, é falar de nossa gente e de seu dia-a-dia, é mostrar que, atrás de uma série de informações estatísticas e dados numéricos, podemos encontrar o rosto do povo brasileiro.

1.4 O *website* do IBGE e o corpus de estudo

O portal do IBGE foi um dos pioneiros entre as páginas institucionais do Brasil, e sua primeira versão data de 1995. Desde o início houve a preocupação de disponibilizar algum conteúdo para possíveis usuários não-falantes do português, e a quantidade de material disponível em outros idiomas variou ao longo do tempo. Uma

versão do *website* em espanhol já esteve disponível, tendo sido suspensa em 2015, por razões administrativas. O *website* em língua inglesa teve início em 1995, e encontra-se plenamente ativo.

Abaixo temos um panorama do *website* do IBGE, com as versões e diferentes *layouts* desde a sua criação:

Figura 2 - Página inicial do *website* do IBGE em 1995

Figura 3 - Página inicial do *website* do IBGE em 1996

Figura 4 - Página inicial do *website* do IBGE em 1998

Figura 7 - Página inicial do website do IBGE em 2007

Figura 8 - Página inicial do website do IBGE em 2017

A página do IBGE na Internet que serviu como base para a pesquisa apresenta informações diversas sobre o Brasil e tem a maior parte de seu conteúdo subdividido em quatro grandes áreas: Indicadores, População, Economia e Geociências⁵, sendo a segunda a que inclui nosso tema de estudo, o Censo Demográfico. Os textos disponíveis *online* e inicialmente considerados para inclusão no corpus deste trabalho apresentam naturezas distintas, em termos de tipo, estrutura e finalidade, a saber:

- Menu lateral - textos informativos e de caráter mais introdutório/genérico;
- Tabelas - títulos e demais legendas para apresentação de resultados numéricos;
- *Releases* - textos jornalísticos publicados em formato de notícia no *website*;
- Questionários - textos com perguntas, para serem lidos em voz alta pelo recenseador;
- *Hotsite* - informações diversas incluindo curiosidades, notícias e FAQ.

Uma breve observação do *website* mostra que nem todo o material produzido e disponibilizado pelo IBGE está traduzido para o inglês. Os *releases* das pesquisas são atualizados regularmente, assim que publicados, mas o mesmo não se dá no caso das “Publicações completas”, que são mais longas e de caráter detalhado, e tampouco dos arquivos com extensão PDF ou *zip*. A versão em inglês do *website* consiste, assim, em uma aproximação do site oficial em português, para que seja possível situar o leitor estrangeiro a respeito das atividades do órgão. É consenso que informações de interesse exclusivamente local, como lançamentos de editais de concurso e contratações, por exemplo, tendem a não ser de interesse do público alóctone. Além desse fato, não há no momento um corpo de tradutores suficiente que possa assumir o volume de traduções de todo o conteúdo.

Embora inicialmente tenha sido considerada a possibilidade de se realizar a pesquisa a partir de todos os tipos de texto disponíveis sobre o Censo, essa proposta foi deixada de lado. Dado o tempo previsto para a conclusão do curso de mestrado,

⁵ A pesquisa baseou-se na versão anterior do *website*, que ficou disponível para consulta até 2017. Atualmente, a página do instituto tem um novo *layout* e apresenta diferenças na organização das informações, com apenas duas grandes áreas, Estatísticas e Geociências, cujo conteúdo foi reagrupado.

dois anos, em média, seria arriscado trabalhar um volume tão grande de informações, em termos de quantidade, e tão diverso, em termos de qualidade. Por esta razão, selecionamos para formar o corpus de estudo apenas dois tipos de texto: *releases* e questionários.

Nossa escolha baseou-se em dois parâmetros: o primeiro foi o de saber quais entre os tipos de textos mencionados estavam melhor disponibilizados e acessíveis à população, e assim selecionamos os *releases* para compor o corpus de estudo. Ao considerarmos a estrutura do *website*, é possível observar que os *releases* sempre aparecem em destaque, uma vez que são notícias atualizáveis regularmente, à medida que os resultados da pesquisa são liberados para acesso público. Além de serem propagados pela própria instituição, os *releases* são previamente repassados à imprensa, com embargo⁶, para que sejam elaboradas matérias para reportagens de telejornais e outros veículos de comunicação. Consideramos, pois, que tais textos tendem a alcançar a sociedade de maneira mais imediata, uma vez que a divulgação de seu conteúdo ultrapassa o âmbito apenas institucional.

Outra característica que torna os *releases* mais acessíveis à sociedade é a sua linguagem simplificada. São textos corridos, de caráter informativo. Há, por parte da instituição, a preocupação de que tais textos sejam de fácil compreensão pelo grande público, sem uso recorrente da terminologia estatística, de termos técnicos, etc. Estes ficam reservados para as seções mais específicas, como, por exemplo, as do tipo 'Metodologia da Pesquisa', não contempladas no presente trabalho.

Para a condução da pesquisa foram selecionados os *releases* relativos ao Censo 2010 disponíveis no *website* do IBGE, em um total de 33, publicados entre os dias 05 de maio de 2010, ou seja, três meses antes do início da coleta de dados, e 16 de março de 2016, quando foi lançada a Grade Estatística⁷ baseada nos resultados do Censo 2010. A única exceção dentro desse período foi a publicação do dia 1º de julho de 2010, referente ao resultado de prova de seleção de profissionais temporários

⁶ Os jornalistas recebem o material com resultados da pesquisa com uma média de 48h de antecedência, com o compromisso de nada revelar antes da divulgação oficial pelo IBGE. Em alguns casos, pode haver uma reunião entre os técnicos do Instituto e os representantes da imprensa, para que sejam dados esclarecimentos que facilitem a elaboração das pautas jornalísticas.

⁷ Divisão do território em células de 200 x 200 m nas áreas urbanas e 1 x 1 km nas áreas rurais, permitindo agregar os dados do Censo 2010 independentemente das divisões político-administrativas.

para atuação no Censo. Tal texto foi desconsiderado para a composição do corpus porque, uma vez que se referia à assunto de interesse estritamente local, não teve um correspondente traduzido para a língua inglesa.

O longo intervalo que separa o primeiro do último release estudado se deu por conta do próprio calendário de apresentação dos resultados do censo. Após a operação censitária propriamente dita, e a finalização da apuração dos resultados – agora facilitada pelo preenchimento digital dos questionários – se dá o início do trabalho de tratamento por temas. Desse modo, as diversas áreas contempladas no Censo, por exemplo, habitação, saúde, religião, cor e raça, serão responsáveis por divulgar os resultados agrupados de forma específica.

O segundo aspecto considerado na seleção do corpus de estudo foi a representatividade do texto quanto à possível presença de características peculiares da população, possivelmente com a ocorrência de marcadores culturais, quesito este que motivou também a própria escolha do Censo como tema para a pesquisa. Tal interesse nos levou a também selecionar para estudo o texto do tipo ‘questionário’, pois este abarca a totalidade das temáticas tratadas na Censo, seja nas perguntas, ou nas opções de resposta oferecidas. Os questionários têm texto resumido e direto, direcionado à obtenção de informações pontuais. Além disso, assim como os *releases*, os questionários também são de certa forma familiares ao público, por serem lidos a cada brasileiro por ocasião da entrevista durante o período dos Censos. Mesmo não estando tão facilmente acessível em sua forma escrita, o questionário é repetido a cada entrevista do Censo, e assim também dialoga com o povo brasileiro em sua totalidade.

Dois tipos de questionário são utilizados durante a operação censitária, atualmente:

- questionário básico: aplicado em todos os domicílios, exceto os selecionados para a amostra dos censos; contempla variáveis como localização e características do domicílio⁸, a lista de moradores, e a relação destes com a pessoa responsável pelo domicílio em questão; tem formato mais resumido;

⁸ Para o IBGE, domicílio é o local estruturalmente separado e independente, que se destina a servir de habitação a uma ou mais pessoas, ou que esteja sendo utilizado como tal. Os critérios essenciais desta definição são os de separação e independência. Entende-se por separação o local de habitação limitado por paredes, muros ou cercas, coberto por um teto, permitindo a uma ou mais pessoas que nele habitam isolar-se das demais, com a finalidade de dormir, preparar e/ou consumir seus alimentos

- questionário da amostra: traz todas as informações contidas no questionário básico, e também outras mais aprofundadas, relativas à educação, trabalho e rendimento, deficiência e fecundidade.

Figura 9 – Questionário básico do Censo 2010

Censo Demográfico 2010 CD 2010 Questionário Básico		MUNICÍPIO: _____ POSTO DE COLETA: _____				
1 IDENTIFICAÇÃO DO DOMICÍLIO						
1.01 UF	1.02 MUNICÍPIO	1.03 DISTRITO	1.04 SUBDISTRITO	1.05 SETOR	1.06 Nº DA QUADRA	1.07 Nº DA FACE
LOCALIDADE: _____				CEP: _____		
LOGRADOURO: _____ (Endereço completo)						
ESPÉCIES DE DOMICÍLIO OCUPADO						
1.08 ☐ 1 - DOMICÍLIO PARTICULAR PERMANENTE OCUPADO ☐ 6 - DOMICÍLIO COLETIVO COM MORADOR ☐ 5 - DOMICÍLIO PARTICULAR IMPROVISADO OCUPADO Siga 1.08						
TIPO						
1.09 ☐ 11 - CASA ☐ 53 - OUTRO (VAGÃO, TRAILER, GRUTA, ETC.) ☐ 12 - CASA DE VILA OU EM CONDOMÍNIO ☐ 61 - ASILO, ORFANATO E SIMILARES COM MORADOR ☐ 13 - APARTAMENTO ☐ 62 - HOTEL, PENSÃO E SIMILARES COM MORADOR ☐ 14 - HABITAÇÃO EM: CASA DE CÔMODOS, CORTIÇO OU CABEÇA DE PORCO ☐ 63 - ALOJAMENTO DE TRABALHADORES COM MORADOR ☐ 15 - OCA OU MALOCA ☐ 64 - PENITENCIÁRIA, PRESÍDIO OU CASA DE DETENÇÃO COM MORADOR ☐ 51 - TENDA OU BARRACA ☐ 65 - OUTRO COM MORADOR ☐ 52 - DENTRO DO ESTABELECIMENTO Obs.: A categoria em negrito só foi disponibilizada em setores de terras indígenas. Se código 1 no quesito 1.08, siga para o quesito 2.01 Se código 5 no quesito 1.08, siga para o quesito 3.01 Se código 6 no quesito 1.08, siga para o quesito 4.01						
2 PARA DOMICÍLIOS PARTICULARES PERMANENTES OCUPADOS CARACTERÍSTICAS DO DOMICÍLIO						
2.01 - ESTE DOMICÍLIO É: ☐ 1 - PRÓPRIO DE ALGUM MORADOR - JÁ PAGO ☐ 4 - CEDIDO POR EMPREGADOR ☐ 2 - PRÓPRIO DE ALGUM MORADOR - AINDA PAGANDO ☐ 5 - CEDIDO DE OUTRA FORMA ☐ 3 - ALUGADO ☐ 6 - OUTRA CONDIÇÃO Siga 2.02						

e proteger-se do meio ambiente, arcando, total ou parcialmente, com suas despesas de alimentação ou moradia. Por independência se entende quando o local de habitação tem acesso direto, permitindo a seus moradores entrar e sair sem necessidade de passar por locais de moradia de outras pessoas. Só caracteriza-se corretamente domicílio quando forem atendidos simultaneamente os critérios de separação e independência.

(<http://censo2010.ibge.gov.br/materiais/guia-do-censo/conceituacao.html>).

Figura 10 – Questionário da amostra do Censo 2010

Censo Demográfico 2010 CD 2010 Questionário da Amostra		MUNICÍPIO: <input type="text"/> POSTO DE COLETA: <input type="text"/>	
1 IDENTIFICAÇÃO DO DOMICÍLIO			
1.01 UF <input type="text"/>	1.02 MUNICÍPIO <input type="text"/>	1.03 DISTRITO <input type="text"/>	1.04 SUBDISTRITO <input type="text"/>
1.05 SETOR <input type="text"/>		1.06 Nº DA QUADRA <input type="text"/>	
1.07 Nº DA FACE <input type="text"/>		1.08 Nº DA FACE <input type="text"/>	
LOCALIDADE: <input type="text"/>		CEP: <input type="text"/>	
LOGRADOURO: (Endereço completo) <input type="text"/>			
ESPÉCIES DE DOMICÍLIO OCUPADO			
1.08 ☐ 1 - DOMICÍLIO PARTICULAR PERMANENTE OCUPADO		☐ 6 - DOMICÍLIO COLETIVO COM MORADOR	
☐ 5 - DOMICÍLIO PARTICULAR IMPROVISADO OCUPADO			
Siga 1.09			
TIPO			
1.09 ☐ 11 - CASA		☐ 53 - OUTRO (VAGÃO, TRAILER, GRUTA, ETC.)	
☐ 12 - CASA DE VILA OU EM CONDOMÍNIO		☐ 61 - ASILO, ORFANATO E SIMILARES COM MORADOR	
☐ 13 - APARTAMENTO		☐ 62 - HOTEL, PENSÃO E SIMILARES COM MORADOR	
☐ 14 - HABITAÇÃO EM: CASA DE CÔMODOS, CORTIÇO OU CABEÇA DE PORCO		☐ 63 - ALOJAMENTO DE TRABALHADORES COM MORADOR	
☐ 15 - OCA OU MALOCA		☐ 64 - PENITENCIÁRIA, PRESÍDIO OU CASA DE DETENÇÃO COM MORADOR	
☐ 51 - TENDA OU BARRACA		☐ 65 - OUTRO MORADOR	
☐ 52 - DENTRO DO ESTABELECIMENTO			
Obs.: A categoria em negrito só foi disponibilizada em setores de terras indígenas. Se código 1 no quesito 1.08, siga para o quesito 2.01 Se código 5 no quesito 1.08, siga para o quesito 3.01 Se código 6 no quesito 1.08, siga para o quesito 4.01			

A decisão sobre o tipo de questionário a ser aplicado em cada domicílio se dá mediante sorteio. No Censo 2010, aproximadamente 89% dos domicílios investigados responderam o questionário básico, e 11%, o modelo mais longo, o chamado questionário da amostra. A partir deste último, os dados foram tratados segundo uma metodologia específica, de modo que os resultados pudessem representar o conjunto total da população. Uma vez que o questionário amostra é mais completo em termos de variedade de informações, e, portanto, mais detalhado quanto ao mapeamento da população brasileira, foi o tipo escolhido para compor nosso corpus de estudo.

É importante destacar que, por conta de mudanças visando à reestruturação do site do IBGE, os questionários em inglês do Censo 2010 atualmente temporariamente inacessíveis pelo *website* do IBGE, podendo ser encontrados no *site* da ONU, apontado como referência neste trabalho. Por esta razão, optou-se por

copiar da mesma fonte também o questionário em português, embora este último ainda esteja disponível na página do Instituto.

Os textos selecionados para o corpus situam-se em dois extremos da operação censitária: os questionários estão no início do processo, ou seja, são o ponto de partida para a obtenção dos resultados propriamente ditos; os *releases* estão na etapa final, isto é, são elaborados quando os dados coletados e já tratados são transformados em informações e números divulgados para a sociedade. Veremos, em seções posteriores desta pesquisa, como tais características podem influenciar as traduções, em especial no que concerne à passagem para a língua inglesa dos marcadores culturais.

2 FUNDAMENTAÇÃO TEÓRICA

O arcabouço teórico-metodológico desta pesquisa concentra-se nos Estudos da Tradução Baseados em Corpus (ETBC) (BAKER, 1993, 1995, 1996; CAMARGO, 2005, 2007) e nos fundamentos da Linguística de Corpus (BERBER SARDINHA, 2000, 2004). Também adotamos a classificação de marcadores culturais entre os domínios ecológico, ideológico, material e social, segundo a proposta de Nida (1945) e sua reformulação por Aubert (1981, 2006), e o conceito de modalidades de tradução (AUBERT, 1998).

Os Estudos da Tradução têm constituído, desde a segunda metade do século XX, uma área que evolui em alcance e complexidade. Como exemplo relevante pertinente a esta trajetória, podemos mencionar os ETBC (BAKER, 1993), que se utilizam de ferramentas tecnológicas para possibilitar a investigação cuidadosa do texto traduzido em um grau de detalhamento provavelmente não alcançável sem o apoio computacional.

É a partir dessa perspectiva, com o olhar voltado ao material produzido por tradutores do IBGE, que conduziremos o estudo aqui proposto, de forma a apresentar um panorama das traduções realizadas e das características registradas por esse “evento comunicativo” (BAKER, 1993), em particular.

Optamos por subdividir a contextualização teórica como se segue: no subitem 2.1, faremos algumas considerações a respeito dos Estudos da Tradução Baseados em Corpus (ETBC) a partir dos Estudos Descritivos da Tradução. Posteriormente, no subitem 2.2, apresentaremos alguns fundamentos da Linguística de Corpus em sua relação com os ETBC, e, por fim, no subitem 2.3, trataremos de características de nosso corpus de estudo, abordando aspectos do texto especializado e da conceituação e tradução de marcadores culturais nesse contexto.

2.1 Os Estudos Descritivos da Tradução e os Estudos da Tradução Baseados em Corpus

Para adequadamente situar esta pesquisa quanto à sua perspectiva teórica, iniciamos a partir do surgimento dos chamados Estudos Descritivos da Tradução, com a semente lançada pelo linguista israelense Itamar Even-Zohar (1978).

Partindo da definição de sistema como uma estrutura de diversos elementos que interagem e se influenciam, a teoria dos polissistemas de Even-Zohar concebe a literatura traduzida como um subsistema pertencente a um sistema muito mais amplo formado pela cultura alvo – e que envolve também os aspectos sociais e históricos, entre outros – formando uma engrenagem continuamente susceptível a mudanças devido a pressões de seus elementos para ocupar o lugar central nesse processo.

O também israelense Gideon Toury (1980, 1995), expande as ideias de Even-Zohar, com quem já havia trabalhado, inclusive. Adotando os pressupostos da teoria dos polissistemas, Toury defende o pensamento de que uma tradução é capaz de influenciar um sistema cultural, mas que também pode ser influenciada por ele, e, considerando tal premissa, desenvolve uma teoria geral da tradução centrada no polo receptor, não mais no texto e na cultura de partida. Segundo o teórico, “a tradução é um tipo de atividade que envolve, inevitavelmente, no mínimo dois idiomas e duas tradições culturais, ou seja, ao menos dois conjuntos de sistemas normativos em cada nível.” (TOURY, 1995, p. 200, tradução nossa⁹)

Para Toury, as traduções são consideradas resultantes de procedimentos gerais que funcionam segundo o polissistema da cultura e língua alvo, e um dos aspectos centrais de sua proposta consiste das normas observáveis em todo trabalho de tradução, ou seja, da existência de leis universais presentes em qualquer processo tradutório. Tais normas pertencem a um modelo do qual também fazem parte a competência e o desempenho, como nos aponta Baker:

“Se considerarmos a competência como um conjunto de todas as opções disponíveis para os tradutores em determinado contexto, e o desempenho como o subconjunto das escolhas de fato feitas pelos tradutores, então as normas consistem em um subconjunto dentro dessas escolhas.” (BAKER, 1993, p.239, tradução nossa¹⁰)

Sendo mensuráveis a partir da comparação entre textos originais (TO) e textos traduzidos (TT), as normas relacionam-se a uma concepção essencialmente descritiva da tradução. Tal abordagem não se resume à análise de um conjunto de

⁹ Tradução nossa para “*Translation is a kind of activity which inevitable involves at least two languages and two cultural traditions, i.e., at least two sets of norm-systems on each level.*”

¹⁰ Tradução nossa para “*If we think of competence as an inventory of all the options that are available to translators in a given context, and performance as the subset of options which are actually selected by translators from this inventory, then norms are a further subset of these options.*”

casos, mas trata da elaboração de princípios gerais que podem ser aplicados em contextos diversos. Avançando em relação à abordagem inicial dos polissistemas, desenvolvidos a partir das traduções literárias, apenas, Toury afirma que, embora nem todas as condições se apliquem simultaneamente, a concepção de normas pode ser estendida a todos os tipos de tradução e a diversos estágios do processo tradutório, subdividindo-se em:

1. Normas preliminares: referem-se à escolha do texto e às estratégias aplicáveis a cada contexto específico;
2. Normas iniciais: tratam da relação entre texto fonte e alvo e das decisões tomadas durante a realização de uma tradução adequada, ou seja, mais próxima às normas que regem o sistema do texto original, ou de uma tradução aceitável, isto é, submetida às normas do sistema da língua alvo;
3. Normas operacionais: referem-se às decisões tomadas especificamente para a composição do texto traduzido, como inclusões, omissões, etc.

O modelo das normas de Gideon Toury foi o ponto de partida para o pensamento de outros estudiosos da tradução, servindo como base para significativas abordagens teóricas de seus contemporâneos, a partir dos desdobramentos dos Estudos Descritivos da Tradução.

É na abordagem interdisciplinar da pesquisadora Mona Baker, elaborada com base nos conceitos propostos por Even-Zohar (1978) e Toury (1980) e nos fundamentos da Linguística de Corpus, que amparamos nossa pesquisa. A premissa fundamental dos Estudos Descritivos da Tradução, de descrever as traduções como ocorrem na realidade, e, a partir dessa observação, chegar a um conjunto de regras gerais, serviram como base para os avanços trazidos posteriormente pelos Estudos da Tradução Baseados em Corpus (BAKER, 1993, 1995, 1996).

Os fundamentos dos ETBC vão ao encontro das necessidades e possibilidades já identificadas no campo dos Estudos Descritivos da Tradução, sendo possível, com sua aplicação, avançar significativamente em direção a uma análise de textos traduzidos em contextos reais.

A relevância do trabalho de Mona Baker no desenvolvimento e consolidação dos ETBC pode ser compreendido em termos do que relata Laviosa:

“Desde suas primeiras propostas com relação ao uso de corpora para o estudo da tradução, Baker tem estado na vanguarda dos estudos baseados em corpus por meio da condução de discussões frutíferas sobre temas de cunho teórico e metodológico e da promoção de um debate aberto e entusiasmado.” (LAVIOSA, 2002, p.23, tradução nossa¹¹)

Entre os aspectos inicialmente apontados por Baker (1993) estão a mudança do foco de atenção do texto fonte para o texto alvo, a passagem de uma visão prescritiva para uma visão descritiva como uma mudança fundamental no caminho dos Estudos da Tradução, e o declínio de uma visão apenas semântica da tradução.

Dando sequência a suas propostas iniciais no universo dos ETBC, Baker (1995) segue em direção à sistematização de uma metodologia de pesquisa em tradução. Apresenta a análise de corpora por meios eletrônicos como uma maneira de perceber diferenças entre a linguagem específica da tradução e a dos textos originalmente escritos em determinada língua, partindo da concepção de corpus, também adotada na presente pesquisa, como “um conjunto de textos naturais (e não de exemplos ou sentenças), em formato eletrônico, passíveis de análise, preferencialmente, de forma automática ou semiautomática (e não manualmente)”. (BAKER, 1995, p.226, tradução nossa¹²).

Ainda com relação à sistematização de pesquisas, define três tipos de corpora passíveis de investigação:

- corpus paralelo: composto de textos originais em uma dada língua e respectivas traduções em outra.
- corpus comparável: composto de dois conjuntos de textos em uma mesma língua, sendo um conjunto de textos originais, e o outro, de textos traduzidos
- corpus multilíngue: composto de dois ou mais corpora monolíngues, sendo cada um em uma língua diferente.

À luz dos pressupostos dos ETBC para a análise de exemplos reais de tradução, Baker (1996) apresenta quatro características que avalia como comuns a todos os textos traduzidos, independentes dos pares de línguas envolvidos, a saber:

¹¹ Tradução nossa para “*Since her first proposals concerning the use of corpora for the study of translation, Baker has been at the forefront of corpus based studies by conducting fruitful discussions on theoretical and methodological issues and keeping the debate open and lively.*”

¹² Tradução nossa para “*any collection of running texts (as opposed to examples/sentences), held in electronic form and analyzable automatically or semi-automatically (rather than manually).*”

- Explicitação: deixar explícita, no texto de chegada, uma informação implícita no texto de partida, seja por questões estilísticas, estruturais ou pragmáticas, o que pode gerar textos traduzidos mais longos que os respectivos originais. Um aspecto que indica a possível ocorrência de explicitação é a maior extensão de um texto traduzido. Outro, é uma maior quantidade de conjunções e locuções adjetivas, uma vez que estas podem servir para explicar informações no TT.

- Simplificação: tornar mais simples a linguagem da tradução, por meio de recursos diversos, em nível lexical ou sintático. Em geral, a simplificação pode ser identificada nos textos traduzidos por meio da resolução de ambiguidades; da quebra de períodos longos e do uso de uma pontuação mais “forte” (por exemplo, a mudança de ponto e vírgula para ponto); e de uma razão forma/item (ver subitem 2.3) mais baixa, o que indica uma menor riqueza vocabular.

- Normalização: incorporar ao texto, de forma exagerada, inclusive, características marcantes da língua de chegada para se adequar aos seus padrões. Como exemplos, podemos mencionar a ocorrências de adições e omissões, mudanças com relação a repetições e o uso diferenciado da pontuação. A análise de resultados obtidos por ferramentas eletrônicas não é um indicativo eficiente para identificar traços de normalização, sendo mais indicado, nesse caso, o cotejo dos textos original e traduzido.

- Estabilização: fazer a tradução de forma a evitar extremos, o que resulta em mais similaridades entre textos traduzidos do que entre textos originais de qualquer tipo.

No nosso caso específico trataremos mais especificamente da investigação de traços dos três primeiros tipos de características da linguagem tradução citadas, ou seja, simplificação, explicitação e normalização. Uma vez que a estabilização só pode ser verificada a partir da comparação entre um maior volume de textos originais e textos traduzidos, não será considerada no presente estudo.

De modo a melhor analisar a tradução dos marcadores culturais selecionados nesta pesquisa, adotaremos como parâmetro para nossa observação a lista de modalidades tradutórias apresentada por Aubert (1998), em uma reformulação das categorizações propostas décadas antes por Vinay e Darbelnet (1958). Segundo o pesquisador, há treze modalidades possíveis:

1. Omissão: ocorre quando não se recupera um dado ou segmento textual.
2. Transcrição: consiste em algarismos, fórmulas e outros que pertençam ao acervo de ambos os idiomas, ou a nenhuma das duas, sendo proveniente de uma terceira língua, ou quando o texto fonte tem uma palavra emprestada na língua meta.
3. Empréstimo: significa um segmento textual do texto fonte reproduzido no texto meta, como topônimos e termos que remetem a realidades específicas
4. Decalque: palavra emprestada da língua fonte, mas submetida a adaptações gráficas e/ou morfológicas que a conformem a convenções da língua fonte.
5. Tradução literal: refere-se à chamada tradução “palavra por palavra”, mantendo-se o número de palavras do segmento, as funções sintáticas e identidade gramatical.
6. Transposição: ocorre quando ao menos um dos critérios que definem a tradução literal não é satisfeito.
7. Explicitação/Implicação: ocorre quando referências implícitas no texto fonte tornam-se explícitas no texto meta, e quando informações explícitas no texto fonte passam a ser referências implícitas.
8. Modulação: observada quando a tradução parece mudar o correspondente semântico de um vocábulo, mas mantém-se o sentido atribuído no contexto, como ocorre, por exemplo, na tradução de expressões idiomáticas e provérbios.
9. Adaptação: ocorre quando há assimilação cultural, sem equivalência perfeita, mas com equivalência parcial de sentido.
10. Tradução intersemiótica: observada quando uma logomarca, figura, brasão ou similar aparece descrita textualmente na língua meta.
11. Erro: quando ocorre mudança de sentido, injustificada e não contextualizada, diferentemente do que ocorre na adaptação.

12. Correção: identificada quando o tradutor “melhora” o texto meta por ocasião de erros, inadequações ou mesmo gafes.

13. Acréscimo: ocorre quando há a inclusão de qualquer segmento textual incluído por conta do tradutor não havendo referência implícita ou explícita no texto original.

2.2. Os Estudos da Tradução Baseados em Corpus e a Linguística de Corpus

O desenvolvimento das abordagens trazidas por Baker com os ETBC relaciona-se diretamente com o uso de ferramentas computacionais para a análise de textos. O sucesso da pesquisa de linguistas como Sinclair (1991) com o uso de compilações de textos em formato eletrônico passa a apontar novos caminhos a serem trilhados também nos Estudos da Tradução. Como abordagem empírica para o estudo da língua, a Linguística de Corpus possibilita, aos pesquisadores em tradução, um novo olhar sobre seu material de estudo a partir do processamento de grandes quantidades de textos. Segundo Berber Sardinha:

“A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador. (BERBER SARDINHA, 2004, p. 3)

A Linguística de Corpus apresenta ao pesquisador evidências para que, a partir delas, possam ser elaboradas leis mais gerais que se apliquem a casos diversos. Para que se possam obter resultados válidos a partir dessa abordagem, entretanto, é preciso atentar para o cumprimento de alguns critérios na compilação do material selecionado para estudo, ou seja do corpus em questão. É preciso que ele seja de linguagem autêntica, tenha a finalidade de estudo linguístico, seja escolhido de forma criteriosa, seja formado por dados legíveis por computador, seja representativo de uma língua ou variedade e, ainda, suficientemente vasto (BERBER SARDINHA, 2002, p. 338). Vale mencionar, entretanto, que além de observar esses critérios, também caberá ao próprio pesquisador compilar ou fazer uso de um corpus que sirva aos objetivos específicos de seu estudo:

Cumpramos ressaltar que tanto os corpora quanto o manuseio de software especializado constituem passos metodológicos importantes, e não os objetivos da pesquisa em si. Há a necessidade de saber interpretar os dados gerados pelas ferramentas de busca à luz de uma abordagem voltada para a exploração de corpora para fins de pesquisa em tradução. (CAMARGO, 2007, p. 97)

Além de uma apropriada delimitação do corpus, o sucesso de uma pesquisa dessa natureza também depende da seleção de uma ferramenta computacional adequada para a obtenção dos resultados. Alguns aspectos interferem nesta escolha: a familiaridade do pesquisador com o universo da informática, a disponibilidade da ferramenta para aquisição, e, também, o grau de dificuldade para sua utilização. Entre as diversas opções atualmente existentes no mercado, optamos por fazer uso, na presente pesquisa, do programa *WordSmith Tools (WST)*, criado em 1996 por Mike Scott, na Universidade de Liverpool, e atualmente em sua sétima versão.

Definido como uma suíte, ou seja, um conjunto de programas integrados, o *WordSmith Tools* apresenta diversas vantagens em relação a seus pares por ser “escrito para o ambiente *Windows*, ou seja, de interface já conhecida, ser também de simples utilização e fácil aquisição” (BERBER SARDINHA, 2009, p.5). Uma breve consulta aos bancos de trabalhos acadêmicos em Estudos da Tradução com o uso de corpora nas universidades brasileiras também valida a escolha do *WordSmith Tools* como ferramenta confiável, uma vez que seu uso é bastante frequente nesses casos.

Os três principais recursos disponíveis no *WordSmith Tools* que têm aplicação no presente estudo, e suas principais funções, são:

- *WordList* - gera listas de palavras, disponibilizadas em ordem alfabética ou de frequência, e, ainda, dados estatísticos dos textos selecionados.
- *KeyWords* - fornece listas das palavras-chave de um texto a partir das listas de palavras do corpus de estudo e de um corpus de referência.
- *Concord* - a partir de palavras inseridas no campo de busca, gera listas de colocados, agrupamentos lexicais, e indica o local de ocorrência da palavra no corpus.

A geração e a interpretação dos resultados baseiam-se nos valores das estatísticas geradas pelo programa quando o texto é processado. Entre os tipos de resultados estatísticos passíveis de serem obtidos, destacamos os que se relacionam aos conceitos abaixo:

- *tokens* ou itens: total de palavras ou ocorrências
- *types* ou formas: total de palavras distintas
- *type/token ratio*: razão forma/item expressa em porcentagem, obtida pela divisão do total de formas pelo total de itens. Indica a riqueza lexical do texto, ou seja, quanto maior a razão, maior o número de palavras diferentes do texto.
- *standardized type/token ratio*: razão forma/item padronizada, calculada para intervalos regulares do texto, com obtenção de um valor médio. É adotada no caso de textos de tamanhos muito diferentes.

Os resultados numéricos gerados pelos programas adotados na Linguística de Corpus para os itens acima sugerem a ocorrência, por exemplo, de determinadas características específicas dos textos traduzidos, conforme visto anteriormente. Por exemplo, a razão *type/token ratio* indica o grau de riqueza vocabular de um texto, e tende a ser maior nos textos originais em comparação às traduções.

Mais observações acerca da utilização do *WordSmith Tools* poderão ser vistas a partir de sua aplicação no tratamento do corpus de estudo, descrita no Capítulo 3, Metodologia.

2.3 O texto especializado e os marcadores culturais

O caráter multidisciplinar da atividade tradutória e, conseqüentemente, dos Estudos da Tradução, permite ao pesquisador realizar sua investigação em áreas diversas do conhecimento, segundo sua necessidade e interesse, entre as quais podemos citar: literatura, formação de tradutores, avaliação de material traduzido, entre outras.

No caso da presente pesquisa, optamos por trazer ao foco de estudo a língua de especialidade, aqui definida como “o conjunto de marcas sintáticas, lexicais, estilísticas e discursivas que tipificam o uso de um código linguístico qualquer em ambiente de interação social centrado em uma determinada atividade humana”. (AUBERT, 2001, p.25)

A tradução especializada, antes comumente identificada pela nomenclatura mais geral “tradução técnica” ou “tradução técnico-científica”, permite que se considere dentro desse campo de estudo textos não necessariamente técnicos, como

os das chamadas ciências duras (Física, Engenharia, e Ciências Exatas em geral), mas aqueles que não pertencem ao universo da língua geral. Podemos citar aqui, por exemplo, os textos jurídicos, institucionais, jornalísticos e os relativos às Ciências Humanas e Sociais.

É inegável que o vocabulário específico pertinente a cada área do conhecimento tem papel fundamental na veiculação dos conteúdos nos chamados textos específicos, sob pena, de caso não devidamente adotado, impedir a compreensão da mensagem ou levar a interpretações equivocadas da informação:

Os textos normalmente têm por objetivo maior transmitir uma informação e, portanto, neles predomina a função referencial. A principal característica desse tipo de texto encontra-se, no entanto, em nível lexical, uma vez que veicula unidades lexicais com conteúdos específicos dos domínios em questão. (BARROS, 2004, p. 44)

Diferentemente do que se acreditou por muitos anos, a tradução de textos dessa natureza não pressupõe apenas que se domine os termos específicos pertencentes a cada área. Expediente bastante adotado, há algumas décadas, era a contratação de profissionais de áreas técnicas e conhecedores da terminologia de determinado ramo científico, mas sem nenhuma formação específica nas áreas de língua ou de tradução. Igualmente comum era a discussão sobre “tradução-técnica x tradução literária”, já criticada com fina ironia na década de 80, por Paulo Rónai:

“O tradutor literário está inclinado a considerar a sua profissão como arte; seu colega técnico, a sua como ciência. Mas nenhuma arte dispensa hoje o aprendizado; o ter nascido com talento não substitui o estudo consciente. Por outro lado, o cientista que, além de criar, deseja comunicar-se com os outros, precisa saber manipular a língua de maneira mais eficiente, isto é, apropriar-se de um estilo, que representa a dimensão estética da linguagem. (RÓNAI, 1983, p.1)

Espera-se que o tradutor de textos especializados disponha de habilidades e capacitação necessárias a tradutores de uma forma geral, e que domine a terminologia pertinente a cada caso em particular. A utilização de dicionários, glossários e listas similares pelos tradutores é prática prevista, mas que se depara com alguns problemas principais: a falta de rigor do profissional na busca pelo termo mais adequado, seja por motivo de imaturidade profissional ou por conta da escassez de tempo, além da interferência de aspectos mercadológicos; a inexistência de material que dê conta dessa busca; e por fim, a própria natureza do texto em questão, principalmente quando os itens a serem buscados apresentam uma carga mais

explícita de conteúdo cultural. Por outro lado, aspectos como o conhecimento da cultura de chegada e do público alvo, além da própria habilidade e criatividade do tradutor, podem contribuir para um resultado final mais satisfatório.

A importância do aspecto cultural na tradução também é reforçada por Soares (2013):

[...] compreendemos que a tradução técnica não se resume apenas ao cuidado com a terminologia e as regras gramaticais dos idiomas envolvidos, visto que parece envolver um olhar bem mais amplo sobre as condições de recepção. Por conseguinte, há uma necessidade de envolvimento maior do tradutor de textos técnicos, no sentido de perceber e refletir também sobre as diferenças culturais entre o polo emissor e o polo receptor, a fim de poder tomar as decisões adequadas durante a atividade tradutória. (SOARES, 2013, p. 2)

De fato, embora os textos especializados apresentem um conjunto vocabular mais relacionado à área a qual se referem, podemos dizer que “relações culturais de toda sorte estarão sempre presentes na tradução de textos especializados (ZAVAGLIA, 2009, p.9).

Os textos sobre o Censo Demográfico podem ser classificados como textos de especialidade, com uma particularidade extra: eles subdividem-se em temáticas diversas, tais como saúde, educação, habitação, etc. Embora não apresentem cunho especificamente regionalista, tais textos podem também serem considerados culturalmente marcados, uma vez que tratam de estruturas sociais e de organização da sociedade que podem não encontrar similar em outras culturas e idiomas:

Outro aspecto a se considerar – e aqui se patenteia a necessidade de uma boa base cultural – é a divergência entre as instituições dos diversos países em decorrência de fatores da evolução histórica. Nem todos os países têm, como nós, ministérios do Planejamento ou da Viação, enquanto nós por enquanto nada temos de parecido com o *Ministère de l'Environnement* da França. À *École Normale Supérieure* de Paris faltam analogias em outras capitais. Os exemplos poderiam ser multiplicados indefinidamente. (RÓNAI, 2012, p. 85)

Observe-se que não trataremos aqui de todos os aspectos culturais possivelmente identificáveis no corpus de estudo, mas concentraremos nossa atenção somente ao nível lexical, especificamente a partir da tradução de determinados vocábulos identificados como marcadores culturais.

Os textos do Censo carregam uma diversidade relativa à realidade brasileira, apresentando não somente dados sobre nosso país, mas particularidades de nosso modo de vida que podem não ser de fácil apreensão para o público estrangeiro. Ao

considerarmos a perspectiva adotada para esse estudo, uma preocupação imediata é refletir sobre a necessária delimitação do que se entende por marcador cultural, uma vez que:

Com efeito, identificação das marcas culturais não constitui uma operação simples e a própria conceituação do que vem a ser uma marca cultural expressa em determinado texto ou ato de enunciação é questão passível de controvérsia. O risco decorrente dessas imprecisões é atribuir a questões de ordem cultural tudo aquilo que não encontra explicação suficientemente convincente no quadro da descrição linguística contrastiva senso estrito. (AUBERT, 2006, p. 24)

Dado o volume do corpus de estudo, a fim de adequar o escopo de nossa análise, faremos um recorte no material selecionado, concentrando-nos nos substantivos e em alguns adjetivos específicos e coincidentes, de relevância para a pesquisa, como, por exemplo, ‘indígenas (s)’:

Ex.1: “Pela primeira vez, o Censo investigará a língua falada pelos indígenas (subst.) residentes no território nacional.”

Ex.2: “Censo investigará línguas indígenas (adj.) pela primeira vez”

Tal recorte mostra ter validade e representatividade, uma vez que as “marcas culturais, insista-se, perpassam todo e qualquer texto, nos mais diversos planos de análise. Na superfície linguística, porém, essas marcas são, efetivamente, mais evidentes no léxico” (Aubert, 2003, p. 154).

A fim de nortear nossa pesquisa, adotaremos a classificação em domínios culturais de Aubert (1981, p. 40-41), em sua reformulação da classificação proposta por Nida (1945). Destacamos, para cada domínio, alguns exemplos obtidos de nosso corpus de estudo.

a) Domínio da cultura ecológica: refere-se a eventos, objetos e seres pertencentes à natureza, seja em estado natural ou aproveitados pelo homem, desde que o conteúdo intrínseco do vocábulo não sugira a ocorrência de ação voluntária do homem. Ex.: igarapé, baixadas

b) Domínio da cultura material: refere-se a objetos criados ou transformados pela atividade humana. Ex.: palafita, pinguela

c) Domínio da cultura social: refere-se ao próprio homem, suas classes e funções sociais, origens, relações hierárquicas, e atividades e eventos que

determinam ou transformam estas relações, incluindo atividades linguísticas.

Ex.: indígenas, municípios

d) Domínio da cultura ideológica: refere-se a crenças, sistemas mitológicos, entidades espirituais, assim como atividades e eventos gerados por essas entidades. Ex.: candomblé, umbanda

Exemplos similares de marcadores culturais em textos especializados são apresentados por Zavaglia (2009), que menciona as diferenças de concepção entre instâncias administrativas e regiões geográficas em países francófonos. Destacamos aqui as considerações a respeito da tradução da palavra ‘prefeito’ para o francês:

Em todos os casos, mesmo que um termo traduza-se por outro (prefeito por *maire*, por exemplo), não refletem exatamente a mesma entidade terminológica, uma vez que as divisões territoriais e administrativas de cada país e as funções de seus governantes são bem distintas (o prefeito brasileiro não celebra, por exemplo, casamentos civis, como o faz o *maire* francês). (ZAVAGLIA, 2009, p.5)

De maneira similar, espera-se que um corpus formado por um conteúdo tão particular quanto o do Censo Demográfico apresente ocorrências dessa natureza quando de sua tradução para o idioma inglês. Especificamente no caso dos estudos populacionais, categoria na qual se insere o Censo, mais do que dados numéricos, tais textos contêm informações sobre um povo e seu modo de vida, em particular. A partir do arcabouço teórico adotado, procuraremos observar detalhadamente a ocorrência de marcadores culturais, descrevendo, segundo os parâmetros de análise propostos, o tratamento dado a eles pela equipe de tradutores do Instituto Brasileiro de Geografia e Estatística, órgão responsável pela realização Censo Demográfico brasileiro.

3 METODOLOGIA

Neste capítulo apresentamos o corpus de estudo da pesquisa (subitem 3.1), bem como os procedimentos metodológicos adotados nesta investigação (subitem 3.2).

3.1 Composição dos corpora

Compilamos, para este estudo, um corpus paralelo formado por textos originais sobre o Censo Demográfico 2010 em português e as respectivas traduções para o inglês, produzidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Ao todo, fazem parte do corpus de estudo quatro subcorpora:

- 33 *releases* jornalísticos em português com as respectivas traduções em inglês, disponíveis no *website* do IBGE;
- 1 questionário do Censo Demográfico 2010 e sua tradução para a língua inglesa, disponíveis no *website* da Divisão de Estatística da ONU.

A data de publicação e os títulos em português e em inglês de cada release estão disponíveis no Apêndice D.

A fim de estudar alguns aspectos do corpus traduzido em relação a textos originalmente escritos em inglês, utilizamos também dois corpora comparáveis compilados a partir de textos provenientes de institutos de estatística internacionais: *Statistics Canada* e *US Census Bureau*.

O primeiro grupo de textos foi obtido das seções *Census in Brief* e *Analytical Document*, do *website* do *Statistics Canada*, que apresenta resultados e comentários sobre a edição de 2011 da pesquisa nacional domiciliar e do censo canadense. Segundo informações disponíveis no *website*, a data mais recente de atualização dos arquivos foi 25 de janeiro de 2017.

O segundo conjunto de textos é proveniente do *website* do *US Census Bureau*, e refere-se à edição americana do Censo 2010. De maneira análoga ao que foi feito no subcorpus do material do IBGE, os textos foram compilados a partir da seção *Newsroom*, e tratam, em linhas gerais, da preparação da operação censitária no país

e dos resultados obtidos, em um total de 65 artigos publicados entre 01 de janeiro de 2010 e 10 de dezembro de 2012.

Para ambos os casos, adotou-se como critério de seleção:

- origem: textos escritos para institutos similares ao IBGE sediados em países cujo inglês é ao menos um dos idiomas oficiais;
- temática: censo ou pesquisa nacional domiciliar
- data: realização em data a mais próxima possível ao Censo 2010; e
- estilo do texto: resultados apresentados como textos corridos (não somente compostos por gráficos, etc.)

Informações adicionais sobre os textos acima referidos encontram-se nos Apêndices E e F, respectivamente.

Para as etapas de obtenção das palavra-chave, utilizamos dois corpora de referência: em português, o Lácio-Ref, corpus de português brasileiro escrito, amplamente empregado em pesquisas dessa natureza; em inglês, escolhemos o *British National Corpus (BNC)*, construído a partir de textos de gêneros diversos em inglês. Estes corpora, formados por textos de linguagem geral, foram utilizados na ferramenta para comparação ao corpus de estudo e subsequente extração das palavras em destaque.

A seguir descrevemos a metodologia utilizada para realizar a compilação e o tratamento dos textos.

3.2 Procedimentos metodológicos

Para o processamento de todos os textos foi adotado o programa *WordSmith Tools*, assim como o cotejo manual, em especial para a extração de exemplos. Todos os procedimentos descritos foram realizados para os textos do tipo *release* e, posteriormente, para os textos do tipo questionário.

3.2.1 Obtenção das listas de palavras

Para darmos início à pesquisa e colocar o material em formato adequado para processamento, os textos foram copiados dos respectivos *websites* e salvos no formato txt, etapa à qual se seguiu a devida correção manual dos erros encontrados (espaçamento, falta de caracteres, etc.).

Inicialmente nos concentramos em reunir informações sobre o corpus de estudo – *releases* e questionário em português e inglês. A primeira ferramenta empregada foi a *WordList*, para a obtenção das listas de palavras e das estatísticas gerais dos TO e TT (Tabela 1). Estando a par destes primeiros resultados, foi possível fazer algumas considerações gerais a respeito dos textos, com relação ao tamanho, frequência de palavras gramaticais, etc.

A seguir ilustramos alguns desses resultados conforme exibidos na tela do programa *WordSmith Tools*.

Figura 11 – Lista de palavras dos *releases* em português, sem edição

N	Word	Freq.	%	Texts	% dispersion	Lemmas	Set
1	DE	3.592	6,51	1	100,00	0,96	
2	E	1.644	2,98	1	100,00	0,98	
3	A	1.482	2,69	1	100,00	0,96	
4	O	1.246	2,26	1	100,00	0,91	
5	EM	1.134	2,05	1	100,00	0,93	
6	DO	871	1,58	1	100,00	0,88	
7	PARA	816	1,48	1	100,00	0,95	
8	QUE	778	1,41	1	100,00	0,93	
9	COM	695	1,26	1	100,00	0,91	
10	DA	687	1,24	1	100,00	0,94	
11	OS	676	1,22	1	100,00	0,95	
12	NO	663	1,20	1	100,00	0,95	
13	DOS	571	1,03	1	100,00	0,91	
14	MAIS	546	0,99	1	100,00	0,87	
15	AS	470	0,85	1	100,00	0,95	

Figura 12 – Lista de palavras dos *releases* em inglês, sem edição

lista de palavras releases completo ingles.lst

N	Word	Freq.	%	Texts	% ispersion	Lemmas	Set
1	THE	4.935	8,91	1	100,00	0,97	
2	OF	2.887	5,21	1	100,00	0,94	
3	IN	2.344	4,23	1	100,00	0,94	
4	AND	1.722	3,11	1	100,00	0,97	
5	TO	1.229	2,22	1	100,00	0,96	
6	WITH	576	1,04	1	100,00	0,86	
7	A	568	1,03	1	100,00	0,93	
8	FROM	487	0,88	1	100,00	0,88	
9	WAS	456	0,82	1	100,00	0,81	
10	POPULATION	451	0,81	1	100,00	0,82	
11	FOR	399	0,72	1	100,00	0,88	
12	CENSUS	361	0,65	1	100,00	0,71	
13	WERE	347	0,63	1	100,00	0,84	
14	PERSONS	341	0,62	1	100,00	0,75	
15	OR	340	0,61	1	100,00	0,86	

Figura 13 – Lista de palavras do questionário em português, sem edição

lista de palavras questionário completo_portugues.lst

N	Word	Freq.	%	Texts	% ispersion	Lemmas	Set
1	DE	162	5,68	1	100,00	0,88	
2	OU	75	2,63	1	100,00	0,87	
3	A	74	2,59	1	100,00	0,77	
4	NÃO	72	2,52	1	100,00	0,83	
5	SIM	59	2,07	1	100,00	0,78	
6	EM	48	1,68	1	100,00	0,81	
7	DO	41	1,44	1	100,00	0,83	
8	ANO	37	1,30	1	100,00	0,63	
9	E	35	1,23	1	100,00	0,76	
10	O	29	1,02	1	100,00	0,68	
11	PARA	28	0,98	1	100,00	0,87	
12	DOMICÍLIO	27	0,95	1	100,00	0,66	
13	QUE	26	0,91	1	100,00	0,75	
14	DA	25	0,88	1	100,00	0,63	
15	JULHO	25	0,88	1	100,00	0,66	

Figura 14 – Lista de palavras do questionário em inglês, sem edição

lista de palavras questionário completo_ingles.lst

N	Word	Freq.	%	Texts	% dispersion	Lemmas	Set
1	IN	76	2,81	1	100,00	0,72	
2	YES	64	2,37	1	100,00	0,80	
3	OR	59	2,18	1	100,00	0,84	
4	THE	56	2,07	1	100,00	0,82	
5	YOU	53	1,96	1	100,00	0,77	
6	NO	50	1,85	1	100,00	0,83	
7	OF	50	1,85	1	100,00	0,74	
8	AND	33	1,22	1	100,00	0,83	
9	UNIT	33	1,22	1	100,00	0,64	
10	TO	32	1,18	1	100,00	0,71	
11	FOR	30	1,11	1	100,00	0,70	
12	HAVE	28	1,04	1	100,00	0,75	
13	THIS	28	1,04	1	100,00	0,74	
14	JULY	26	0,96	1	100,00	0,64	
15	HOUSING	25	0,93	1	100,00	0,55	

A seguir fizemos a edição das *WordList* inicialmente obtidas, deixando nas listas apenas o nosso vocabulário de interesse, ou seja, os substantivos e as palavras que pudessem ser classificadas como adjetivo ou substantivo. Para esta identificação fizemos uso da ferramenta *Concord*, que também permite melhor visualização dos colocados presentes no texto. Apresentamos a seguir um dos casos em que a ferramenta foi de grande utilidade para definir o contexto de ocorrência de um vocábulo:

Figura 15 – Lista de concordância para ‘aglomerados’ a partir dos *releases*

releases completo portugues_aglomerados.cnc

N	Concordance	Word #	Sent	Sent	Para	Para	Para	Para	Para	Para	File	Date	%
1	Norte, os trabalhadores informais vivendo em aglomerados chegavam a 36,0%, enquanto	52.573	1.8	53%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
2	capital onde essa proporção era mais alta nos aglomerados (48,3%, contra 29,3% nas	52.595	1.8	70%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
3	para o trabalho era menor para moradores de aglomerados no Rio de Janeiro O tempo	52.613	1.8	31%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
4	Sul, onde 19,1% da população residente em aglomerados subnormais e 15,1% dos	52.550	1.8	58%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
5	a maior proporção de pessoas residentes em aglomerados com curso superior (2,8%),	52.481	1.8	52%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
6	tinha 48,3% dos trabalhadores residentes em aglomerados não tinham carteira assinada; no	52.502	1.8	25%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
7	relações de trabalho era maior nas áreas de aglomerados (27,8%) do que no restante da	52.524	1.8	80%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
8	variava muito entre as pessoas que moram em aglomerados subnormais e as outras áreas da	52.635	1.8	83%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
9	pontos percentuais: 28,4% dos domicílios em aglomerados subnormais estão nesta faixa de	52.878	1.8	57%	0	97%	0	97%	releases	comple	2018-abr-19 00:		96%
10	enquanto 0,9% da população residente em aglomerados se enquadrava nesta faixa de	52.916	1.8	67%	0	97%	0	97%	releases	comple	2018-abr-19 00:		96%
11	1,2% e 20,8%, respectivamente. Moradores de aglomerados têm menos carros, máquinas de	52.950	1.8	8%	0	97%	0	97%	releases	comple	2018-abr-19 00:		96%
12	per capita, 31,6% da população residente em aglomerados subnormais estava na faixa até	52.821	1.8	80%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
13	Na capital fluminense, 21,9% da população de aglomerados levava mais de uma hora por dia	52.712	1.8	31%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
14	: 37,0% da população residente em aglomerados leva mais que uma hora para	52.774	1.8	38%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%
15	áreas da cidade. 31,6% da população dos aglomerados tinham renda de até meio salário	52.801	1.8	63%	0	96%	0	96%	releases	comple	2018-abr-19 00:		96%

A lista a seguir apresenta o vocabulário de interesse para a pesquisa propriamente dita:

Figura 16 – Lista de palavras dos *releases* em português, editada

N	Word	Freq.	%	Texts	% ispersion	Lemmas	Set
1	ANOS	378	0,68	1	100,00	0,77	
2	POPULAÇÃO	356	0,65	1	100,00	0,85	
3	DOMICÍLIOS	338	0,61	1	100,00	0,94	
4	CENSO	332	0,60	1	100,00	0,74	
5	PESSOAS	315	0,57	1	100,00	0,85	
6	IBGE	196	0,36	1	100,00	0,59	
7	RENDIMENTO	183	0,33	1	100,00	0,79	
8	MULHERES	167	0,30	1	100,00	0,70	
9	MUNICÍPIOS	165	0,30	1	100,00	0,64	
10	REGIÃO	165	0,30	1	100,00	0,83	
11	TRABALHO	165	0,30	1	100,00	0,75	
12	INDÍGENAS	163	0,30	1	100,00	0,09	
13	IDADE	159	0,29	1	100,00	0,75	
14	PAÍS	152	0,28	1	100,00	0,85	
15	ÁREAS	138	0,25	1	100,00	0,64	

Abaixo observamos uma lista equivalente, baseada nos textos traduzidos. Como a anterior, os itens parecem em ordem decrescente de frequência:

Figura 17 – Lista de palavras dos *releases* em inglês, editada

N	Word	Freq.	%	Texts	% ispersion	Lemmas	Set
1	POPULATION	451	0,81	1	100,00	0,82	
2	CENSUS	361	0,65	1	100,00	0,71	
3	PERSONS	341	0,62	1	100,00	0,75	
4	YEARS	255	0,46	1	100,00	0,76	
5	AREAS	230	0,42	1	100,00	0,69	
6	INCOME	219	0,40	1	100,00	0,82	
7	HOUSEHOLDS	209	0,38	1	100,00	0,76	
8	AGE	198	0,36	1	100,00	0,66	
9	IBGE	197	0,36	1	100,00	0,57	
10	PERCENTAGE	192	0,35	1	100,00	0,76	
11	UNITS	192	0,35	1	100,00	0,75	
12	WOMEN	183	0,33	1	100,00	0,67	
13	MUNICIPALITIES	173	0,31	1	100,00	0,67	
14	REGION	166	0,30	1	100,00	0,80	
15	HOUSEHOLD	145	0,26	1	100,00	0,83	

De forma análoga ao que se vê na Figura 16, apresentamos aqui uma amostra dos dados obtidos para o questionário:

Figura 18 – Lista de palavras do questionário em português, editada

lista de palavras questionário editado portugues.lst							
File Edit View Compute Settings Windows Help							
N	Word	Freq.	%	Texts	%ispersion	Lemmas	Set
1	ANO	37	1,29	1	100,00	0,63	
2	DOMICÍLIO	27	0,94	1	100,00	0,66	
3	JULHO	25	0,87	1	100,00	0,66	
4	TRABALHO	22	0,77	1	100,00	0,33	
5	PAÍS	21	0,73	1	100,00	0,49	
6	MUNICÍPIO	19	0,66	1	100,00	0,45	
7	SÉRIE	17	0,59	1	100,00	0,00	
8	PESSOA	16	0,56	1	100,00	0,59	
9	DIFICULDADE	12	0,42	1	100,00	0,24	
10	FEDERAÇÃO	12	0,42	1	100,00	0,18	
11	MORADOR	12	0,42	1	100,00	0,72	
12	UNIDADE	12	0,42	1	100,00	0,18	
13	FILHO	10	0,35	1	100,00	0,20	
14	MÊS	9	0,31	1	100,00	0,22	
15	NOME	9	0,31	1	100,00	0,46	

E o mesmo para o questionário traduzido:

Figura 19 – Lista de palavras do questionário em inglês, editada

lista de palavras questionário editado_ingles.lst							
File Edit View Compute Settings Windows Help							
N	Word	Freq.	%	Texts	%ispersion	Lemmas	Set
1	UNIT	33	1,21	1	100,00	0,64	
2	JULY	26	0,95	1	100,00	0,64	
3	HOUSING	25	0,92	1	100,00	0,55	
4	COUNTRY	20	0,73	1	100,00	0,49	
5	MUNICIPALITY	19	0,70	1	100,00	0,56	
6	EDUCATION	17	0,62	1	100,00	0,07	
7	YEAR	17	0,62	1	100,00	0,48	
8	FEDERATION	16	0,59	1	100,00	0,28	
9	PERSON	15	0,55	1	100,00	0,28	
10	RESIDENTS	12	0,44	1	100,00	0,69	
11	JOB	11	0,40	1	100,00	0,31	
12	RESIDENT	11	0,40	1	100,00	0,67	
13	WORK	11	0,40	1	100,00	0,49	
14	BIRTH	9	0,33	1	100,00	0,46	
15	DIFFICULTY	9	0,33	1	100,00	0,12	

Para analisar os aspectos mais gerais do corpus de estudo, optamos por fazer um recorte das trinta ocorrências mais frequentes das *WordList*. Esse número representa aproximadamente 5% do número de palavras total e foi posteriormente mantido para a análise dos corpora comparáveis. Embora possa parecer uma porcentagem baixa, esta amostra é bastante relevante em termos qualitativos, considerando-se que, no caso dos *releases* em português, por exemplo, as palavras que ocorrem apenas uma vez totalizam 42% do texto. Assim, os primeiros 5% constituem uma amostra considerável do material mais relevante no corpus.

3.2.2 Listagem das palavras-chave

Para chegarmos à lista com as palavras-chave, a partir da qual é possível identificar mais facilmente os possíveis marcadores culturais, foi empregada a ferramenta *KeyWords*.

A partir da comparação entre as palavras do TO e o corpus de referência obtivemos um conjunto de palavras-chave, para ambos os tipos de texto (*releases* e questionário).

Figura 20 – Palavras-chave dos *releases* em português

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Log_L	Log_R	P Lemmas	Set
1	DOMICÍLIOS	338	0,61	1	14		2.464,23	10,07	0,0000000000	
2	CENSO	332	0,60	1	45		2.261,81	8,36	0,0000000000	
3	POPULAÇÃO	356	0,65	1	811	0,03	1.319,20	4,29	0,0000000000	
4	IBGE	196	0,36	1	103		1.116,35	6,41	0,0000000000	
5	RENDIMENTO	183	0,33	1	105		1.024,39	6,28	0,0000000000	
6	INDÍGENAS	163	0,30	1	83		934,00	6,45	0,0000000000	
7	AGLOMERADOS	122	0,22	1	9		866,55	9,24	0,0000000000	
8	PESSOAS	315	0,57	1	1.885	0,08	682,32	2,90	0,0000000000	
9	IDADE	159	0,29	1	333	0,01	609,92	4,41	0,0000000000	
10	FECUNDIDADE	77	0,14	1	4		556,37	9,74	0,0000000000	
11	NORTE	130	0,24	1	224		537,28	4,69	0,0000000000	
12	MULHERES	167	0,30	1	585	0,02	504,97	3,67	0,0000000000	
13	ANOS	378	0,68	1	3.959	0,16	495,91	2,09	0,0000000000	
14	RECENSEADORES	64	0,12	1	0		488,77	139,25	0,0000000000	
15	SUDESTE	95	0,17	1	88		476,01	5,59	0,0000000000	

Figura 21 – Palavras-chave do questionário em português

palavras chave questionário português.kws										
File Edit View Compute Settings Windows Help										
N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Log_L	Log_R	P	Set
1	DOMICÍLIO	27	0,94	1	6		329,60	11,81	0,0000000000	
2	JULHO	25	0,87	1	169		185,49	6,88	0,0000000000	
3	MUNICÍPIO	19	0,66	1	49		173,51	8,27	0,0000000000	
4	ANO	37	1,29	1	1.135	0,05	168,85	4,70	0,0000000000	
5	MORADOR	12	0,42	1	15		123,33	9,32	0,0000000000	
6	FEDERAÇÃO	12	0,42	1	31		109,55	8,27	0,0000000000	
7	SÉRIE	17	0,59	1	491	0,02	79,52	4,79	0,0000000000	
8	PAÍS	21	0,73	1	1.080	0,05	75,50	3,96	0,0000000000	
9	DIFICULDADE	12	0,42	1	212		67,34	5,50	0,0000000000	
10	CÔNJUGE	5	0,17	1	0		66,83	139,84	0,0000000000	
11	PESSOA	16	0,56	1	627	0,03	65,63	4,35	0,0000000000	
12	RENDIMENTO	7	0,24	1	20		62,71	8,13	0,0000000000	
13	DOMICÍLIOS	5	0,17	1	1		61,43	11,96	0,0000000000	
14	CRECHE	6	0,21	1	9		60,03	9,05	0,0000000000	
15	UNIDADE	12	0,42	1	371	0,02	54,59	4,69	0,0000000000	

Por meio da observação das palavras-chave selecionamos os vocábulos que poderiam ser considerados marcadores culturais. O critério de seleção dos marcadores a serem analisados na pesquisa foi, em primeiro lugar, o grau de chavicidade – ou seja, de especificidade das palavras – apresentado pela própria ferramenta *KeyWords*. Em segundo lugar, adotamos o critério de possível pertença a um dos quatro domínios culturais redefinidos por Aubert (1981), conforme mencionado no Capítulo 2.

Estabelecemos, para fins desta pesquisa, um limite de seis marcadores culturais por domínio, resultando em um número máximo de 24 marcadores a serem analisados. Para completar a lista, primeiramente foram selecionados todos os marcadores identificados dentre as palavras-chave, e, havendo ainda lacunas, recorreremos às listas de palavras editadas para identificar outros possíveis marcadores culturais.

3.2.3 Levantamento e classificação dos marcadores culturais

Os marcadores culturais selecionados a partir das palavras-chave e listas de palavras foram devidamente classificados entre os quatro domínios culturais – ecológico, ideológico, material e social. De modo a justificar e validar a escolhas dos referidos marcadores, apresentamos, para cada um, observações acerca do seu significado e utilização no contexto específico do censo. Para fins de conceituação, recorreremos ao dicionário *Webster Online*.

O passo seguinte foi contabilizar as ocorrências de cada marcador selecionado e calcular a porcentagem das opções tradutórias adotadas. Assim, foi possível exibir um panorama da distribuição percentual para a tradução dos marcadores em ambos os tipos de texto.

3.2.4 Análise das traduções

O primeiro nível de análise referiu-se apenas às diferentes opções tradutórias adotadas ao longo dos textos. A comparação entre os textos originais e traduzidos para a observação de tais correspondências foi realizada por meio do recurso *Aligner*, que, como o próprio nome indica, alinha os períodos de forma que se possa visualizar par a par as escolhas feitas pela equipe de tradução. Em muitos casos foi necessário o cotejo dos textos, pois a ferramenta não se mostrou de uso tão rápido e eficiente. Os resultados incluem comentários e exemplos acerca de cada marcador cultural, englobando as diversas soluções tradutórias e ressaltando as diferenças entre aspectos observados nos dois textos estudados: *releases* e questionários.

De modo a enriquecer a discussão, optamos por submeter alguns dos textos em que ocorrem os marcadores culturais selecionados à leitura de um nativo da língua inglesa, o professor britânico Donald Hunter. Assim, em alguns casos, apresentaremos, além das traduções publicadas no *website*, uma solução tradutória sugerida e comentários deste falante nativo acerca das traduções tratadas. Vale ressaltar, no entanto, que as equipes de tradução do IBGE são responsáveis tanto pela revisão quanto pela tradução de seu próprio material, e que no seu dia-a-dia de trabalho não dispõem de um serviço deste tipo.

O segundo nível de análise baseou-se nas características da tradução (BAKER, 1996). A partir da observação de excertos de texto original que continham marcadores, e da respectiva versão traduzida, foram compilados exemplos de explicitação, normalização e simplificação.

Por último, os textos do tipo *release* e questionário foram comparados aos textos dos corpora originalmente escritos em inglês, provenientes do *Statistics Canada* e do *US Census Bureau*. Inicialmente foram obtidas as listas de palavras gerais e, a partir destas, as respectivas listas de palavras-chave. Neste caso adotamos o mesmo corpus de referência utilizado anteriormente, o *BNC*.

Figura 22 – Palavras-chave dos textos do *Statistics Canada*

statcan wordlist2.kws										
File Edit View Compute Settings Windows Help										
N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Log_L	Log_R	P Lemmas	Set
1	POPULATION	783	1,20	1	12.99	0,01	5.487,28	6,52	0,0000000000	
2	CENSUS	493	0,76	1	1.037		5.306,98	9,50	0,0000000000	
3	CANADA	466	0,72	1	2.695		4.192,63	8,04	0,0000000000	
4	INCOME	449	0,69	1	12.00	0,01	2.734,30	5,84	0,0000000000	
5	ABORIGINAL	197	0,30	1	212		2.322,47	10,47	0,0000000000	
6	HOUSEHOLDS	246	0,38	1	1.772		2.113,31	7,73	0,0000000000	
7	HOUSEHOLD	195	0,30	1	3.837		1.302,61	6,28	0,0000000000	
8	PROPORTION	219	0,34	1	6.543		1.286,63	5,68	0,0000000000	
9	PROVINCES	140	0,21	1	1.114		1.176,66	7,58	0,0000000000	
10	STATISTICS	173	0,27	1	3.197		1.176,54	6,37	0,0000000000	
11	AGE	285	0,44	1	21.48	0,02	1.169,57	4,34	0,0000000000	
12	PERCENTAGE	166	0,25	1	2.806		1.157,39	6,50	0,0000000000	
13	IMMIGRANTS	129	0,20	1	831		1.134,96	7,89	0,0000000000	
14	INUIT	93	0,14	1	75		1.132,81	10,89	0,0000000000	
15	QUEBEC	103	0,16	1	285		1.061,61	9,11	0,0000000000	
16	ALBERTA	89	0,14	1	90		1.057,00	10,56	0,0000000000	
17	FAMILIES	196	0,30	1	8.222		1.023,42	5,19	0,0000000000	
18	CHILDREN	326	0,50	1	41.33	0,04	1.022,44	3,59	0,0000000000	
19	TERRITORIES	126	0,19	1	1.377		983,54	7,13	0,0000000000	
20	COUPLES	126	0,19	1	1.547		955,62	6,96	0,0000000000	
21	ONTARIO	94	0,14	1	328		931,14	8,77	0,0000000000	

Figura 23 – Palavras-chave dos textos do *US Census Bureau*

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Log_L	Log_R	P	Lemmas	Set
1	CENSUS	1.114	2,67	65	11		7.084,57	11,27	0,0000000000		
2	POPULATION	557	1,34	40	136	0,01	2.928,33	6,64	0,0000000000		
3	BUREAU	299	0,72	65	42		1.683,34	7,44	0,0000000000		
4	HOUSEHOLDS	112	0,27	28	2		704,64	10,42	0,0000000000		
5	RACE	153	0,37	17	103	0,01	653,09	5,18	0,0000000000		
6	MAIL	129	0,31	27	47		634,12	6,07	0,0000000000		
7	HISPANIC	79	0,19	12	0		511,12	139,96	0,0000000000		
8	STATES	217	0,52	53	598	0,06	507,40	3,15	0,0000000000		
9	ASIAN	82	0,20	9	10		468,07	7,64	0,0000000000		
10	AREAS	124	0,30	22	234	0,02	359,11	3,69	0,0000000000		
11	GROVES	60	0,14	32	4		358,59	8,52	0,0000000000		
12	PARTICIPATION	78	0,19	26	41		354,67	5,54	0,0000000000		
13	BLACK	110	0,26	13	207	0,02	319,01	3,70	0,0000000000		
14	NATIVE	73	0,18	11	46		317,20	5,28	0,0000000000		
15	NATION	100	0,24	41	169	0,02	305,54	3,85	0,0000000000		
16	DATA	100	0,24	44	170	0,02	304,69	3,84	0,0000000000		
17	PERCENTAGE	68	0,16	24	45		291,63	5,20	0,0000000000		
18	PEOPLE	182	0,44	42	856	0,08	282,48	2,38	0,0000000000		
19	RESPONDENTS	52	0,12	32	13		272,42	6,61	0,0000000000		
20	ALASKA	55	0,13	17	22		265,47	5,93	0,0000000000		

Havendo novamente estabelecido como 30 ocorrências o limite de corte, listamos as palavras-chave dos textos do IBGE e dos dois institutos internacionais, de modo a identificar aspectos coincidentes e contrastantes a partir destas ocorrências.

O mesmo procedimento foi seguido para os questionários dos três institutos, com uma modificação. Uma vez que o questionário do *Statistics Canada* gerou uma lista de somente 15 palavras-chave, este foi o valor de corte adotado da segunda vez.

Todas as análises e conclusões baseiam-se nos resultados gerados pelo programa *WordSmith Tools*, tendo como ponto de partida a identificação dos marcadores culturais e a observação das respectivas traduções.

A apresentação e discussão dos resultados encontra-se na seção a seguir.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Neste capítulo apresentamos os resultados obtidos na análise das traduções de marcadores culturais para a língua inglesa a partir de 33 *releases* sobre o Censo Demográfico 2010, publicados no *website* do IBGE entre 05 de maio de 2010 e 16 de março de 2016, e do questionário da amostra (formato mais completo) aplicado para a obtenção de dados dessa pesquisa populacional. O presente capítulo subdivide-se nas seguintes seções: 4.1 Estatísticas gerais; 4.2 Observações acerca das listas de palavras em português e em inglês; 4.3 Palavras-chave dos textos originais e traduzidos; 4.4 Seleção, significado e classificação dos marcadores culturais; 4.5 Análise das traduções dos marcadores culturais; 4.6 Análise de características da tradução; 4.7 Comparação com textos de inglês nativo

4.1 Estatísticas gerais

Os primeiros resultados obtidos por meio do programa *WordSmith Tools* referem-se às estatísticas mais genéricas dos textos apresentadas pela ferramenta *WordList*, conforme detalhado nas tabelas abaixo:

Tabela 1 – Estatísticas gerais dos corpora de estudo obtidas pela ferramenta *WordList*

	Itens (Tokens)	Formas (Types)	Razão forma/item (Type/token ratio)	Razão forma/item padronizada (Standardized type/token ratio)
<i>Releases</i> em português	50.148	4.373	8,72	33,05
<i>Releases</i> em inglês	50.389	3.488	6,92	30,24
Questionário em português	2.308	623	26,99	28,85
Questionário em inglês	2.163	574	26,54	28,70

Tabela 2 - Diferenças percentuais entre as estatísticas gerais das traduções em relação aos textos originais

	Diferenças entre itens	Diferenças entre formas	Diferenças entre razão forma- item padronizada
<i>Releases</i>	+ 0.48%	- 26.33%	-8.5%
Questionário	- 6.28%	-7.87%	-0.51%

Uma primeira observação relevante refere-se ao tamanho dos textos, conforme ilustrado na coluna 'Itens', da Tabela 1. Confirmando o que é geralmente esperado no caso de traduções, nos *releases* o TT é maior do que o original, sendo 50,148 itens no TO e 50.389 no TT. Esta diferença é mínima, no entanto, apenas 0.48% (Tabela 2). Considerando o questionário, os resultados contrariam o que se prevê na teoria: temos 2.308 itens no TO e 2.163 no TT, ou seja, conforme explicitado na Tabela 2, o questionário traduzido apresenta diminuição de 6.28% em seu tamanho. Essas porcentagens – de quase equiparação e de redução – podem indicar uma baixa incidência de explicitações nas traduções em inglês, refletindo a possível opção dos tradutores de adotar uma dada solução tradutória sem qualquer explicação adicional para situar o leitor estrangeiro acerca de tais significados no contexto original.

A taxa de ocorrência de formas (*types*) em relação ao número total de itens (*tokens*), medida por meio da razão forma/item (*type/token ratio*) foi maior nos textos em português, indicando maior riqueza vocabular dos originais em relação às traduções, conforme o que indica a teoria. Quanto maior a razão forma/item, mais variado é o texto em termos de léxico, ocorrendo o oposto à medida que essa razão diminui.

Como os subcorpora apresentam tamanhos diferentes, consideramos também a razão forma/item padronizada, *standardized type/token ratio*, na última coluna à direita da tabela. Os valores 33,05 dos *releases* em português versus 30,24 em inglês, e 28,85, do questionário em português versus 28,70, do questionário em inglês, evidenciam que os corpora analisados seguem padrões geralmente observados, de menor riqueza vocabular dos textos traduzidos, possivelmente indicando a ocorrência de simplificações. Nos dois casos, entretanto houve relativamente pouca diferença entre os textos originais e traduzidos, e uma variação ínfima no caso do questionário, em especial. Tal aspecto sugere uma tradução mais caracterizada pelo aspecto literal e pela manutenção de estruturas similares à do texto original.

Vale observar que, ao analisarmos a coluna *types*, podemos notar significativa diferença entre os *releases* e o questionário. O questionário apresenta 7.87% menos formas quando traduzido para o inglês, ao passo que, no caso dos *releases*, esta diferença entre TO e TT sobe para 26.33%, ou seja, quase quatro vezes mais. Também observamos que a diferença entre a razão forma/item padronizada do TO e do TT foi de 8.5%, no caso dos *releases*, e de apenas 0.51%, no questionário. Esses valores nos sugerem que, entre os dois tipos de textos traduzidos, o questionário

apresenta perda de riqueza vocabular significativamente menor do que os *releases*. Uma possível explicação para esse fato é que, sendo mais longos, os *releases* podem se tornar também mais repetitivos, o que pode ter levado os tradutores a fazer as escolhas tradutórias similares em diversos trechos do texto. O questionário, ao contrário, é mais objetivo e trata de aspectos muito específicos em cada pergunta que o compõe, o que impediria a repetição dos mesmos vocábulos na tradução.

De forma geral, tais valores sugerem que o formato intrínseco ao questionário dificulta um afastamento significativo do original em português. Ao contrário, no texto corrido dos *releases* pode ter havido duas ocorrências distintas: um maior número de simplificações e, ainda, a incidência de omissões nos trechos de maior dificuldade ou redundância.

4.2 Observações acerca das listas de palavras em português e em inglês

As listas de palavras dos textos originais e traduzidos apresentaram significativa quantidade de itens gramaticais, como preposições, numerais, conjunções, entre outras que designam diferentes classes. As tabelas 3 e 4 (referentes aos *releases* e ao questionário, respectivamente) mostram as 30 maiores frequências, em ordem decrescente, obtidas no caso das listas completas, ou seja, sem edição:

Tabela 3 – Trinta maiores ocorrências das listas de palavras, sem edição, dos *releases* em português (TO) e em inglês (TT)

N	Palavra (TO)	Frequência (TO)		Palavra (TT)	Frequência (TT)
1	de	3.592		the	4.935
2	e	1.644		of	2.887
3	a	1.482		in	2.344
4	o	1.246		and	1.722
5	em	1.134		to	1.229
6	do	871		with	576
7	para	816		a	568
8	que	778		from	487
9	com	695		was	456
10	da	687		population	451
11	os	676		for	399
12	no	663		census	361
13	dos	571		were	347
14	mais	546		persons	341
15	as	470		or	340
16	na	403		by	312
17	ou	396		that	290
18	das	379		years	255
19	anos	378		is	252
20	população	356		than	246
21	domicílios	338		as	244
22	censo	332		more	231
23	pessoas	315		areas	230
24	por	299		income	219
25	entre	292		this	216
26	foi	259		among	209
27	um	241		households	209
28	maior	228		million	208
29	não	219		thousand	199
30	mil	204		age	198

O limite de trinta foi escolhido como fator de corte para exibirmos aproximadamente 5% do total de ocorrências. Como esperado, encabeçam a lista, em ambos os idiomas, artigos, conjunções, preposições e pronomes, entre outros.

Desta parte destacamos alguns aspectos: do total de trinta vocábulos apresentados, aparecem, em português, cinco substantivos de significado bastante relevante dentro do tema Censo Demográfico: ‘anos’, ‘população’, ‘domicílios’, ‘censo’ e ‘pessoas’. Em inglês, há sete desses substantivos: ‘*population*’, ‘*census*’, ‘*years*’, ‘*areas*’, ‘*income*’, ‘*households*’, ‘*age*’. Para a seleção em língua portuguesa a primeira palavra de conteúdo semântico relevante aparece na 20ª posição: ‘população’, com 356 ocorrências. No caso do texto traduzido, a primeira palavra de conteúdo similar é seu provável correspondente em inglês, ‘*population*’, que surge na décima posição e com um número significativamente maior de ocorrências, 451.

A disparidade de posições na lista pode resultar da própria diferença entre a natureza das palavras gramaticais nos dois idiomas, uma vez que o inglês não permite a flexão de número ou gênero no caso das preposições, por exemplo. Assim, enquanto a lista do texto original traz 'de', 'do', 'da', 'dos', 'das', a do texto traduzido apresenta duas traduções mais imediatas 'of' e 'from'. A isso some-se também a possibilidade do uso do genitivo ('s), em substituição à preposição 'de', do português, e a omissão de preposições em locuções adjetivas, entre outras particularidades da gramática do idioma inglês que podem explicar a aparente discrepância entre as listas sem edição.

Análise similar é apresentada para o questionário, do qual também retiramos uma amostra das trinta primeiras ocorrências nos textos original e traduzido.

Tabela 4 – Trinta maiores ocorrências das listas de palavras, sem edição, dos questionários em português (TO) e em inglês (TT)

N	Palavra (TO)	Frequência (TO)		Palavra (TT)	Frequência (TT)
1	de	162		<i>in</i>	76
2	ou	75		<i>yes</i>	64
3	a	74		<i>or</i>	59
4	não	72		<i>the</i>	56
5	sim	59		<i>you</i>	53
6	em	48		<i>no</i>	50
7	do	41		<i>of</i>	50
8	ano	37		<i>and</i>	33
9	e	35		<i>unit</i>	33
10	o	29		<i>to</i>	32
11	para	28		<i>for</i>	30
12	domicílio	27		<i>have</i>	28
13	que	26		<i>this</i>	28
14	da	25		<i>July</i>	26
15	julho	25		<i>housing</i>	25
16	qual	25		<i>is</i>	23
17	é	24		<i>what</i>	21
18	trabalho	22		<i>a</i>	20
19	país	21		<i>country</i>	20
20	município	19		<i>do</i>	19
21	estrangeiro	18		<i>municipality</i>	19
22	série	17		<i>your</i>	18
23	mais	16		<i>education</i>	17
24	pessoa	16		<i>year</i>	17
25	na	15		<i>federation</i>	16
26	no	15		<i>foreign</i>	16
27	outro	13		<i>person</i>	15
28	dificuldade	12		<i>at</i>	14
29	federação	12		<i>from</i>	14
30	morador	12		<i>did</i>	13

No caso do questionário, diferentemente do exemplo anterior, há maior número de palavras semânticas no texto em português, com os substantivos: ‘ano’, ‘domicílio’, ‘julho’, ‘trabalho’, ‘país’, ‘município’, ‘série’, ‘pessoa’, ‘dificuldade’, ‘federação’, ‘morador’, e ainda ‘estrangeiro’, que poderia, inicialmente, ser classificada tanto com substantivo quanto como adjetivo.

No texto em inglês as ocorrências trazem possíveis correspondentes tradutórios para as palavras semânticas em português, por exemplo ‘*July*’/‘julho’, ‘*federation*’/‘federação’, ‘*pessoa*’/‘*person*’. Entre as palavras gramaticais, ressalta-se aqui a significativa presença do pronome ‘*you*’, no TT, sugerindo que as perguntas em inglês fizeram provável uso da segunda pessoa do singular, ou seja, foram dirigidas ao interlocutor do agente recenseador e, portanto, respondente do questionário, como se espera em ocasiões do tipo. Em português, ao contrário, a pergunta não requer o pronome em questão, e poderia ser perfeitamente entendida ao manter-se oculto o sujeito da oração, conforme destacado nos exemplos abaixo:

TO: 6.08 - Qual é a sua etnia ou o povo a que pertence?

TT: 6.08 - *What ethnic group or people do you belong to?*

TO: 6.11 - Fala português no domicílio?

TT: 6.11 - *Do you speak Portuguese in the housing unit?*

Dando prosseguimento a esta etapa, realizou-se a eliminação das palavras gramaticais, com o intuito de filtrar das listas os principais itens de interesse nesta pesquisa, em especial os substantivos. Optamos por manter também nas listas as palavras que carregam significado e podem fazer o papel tanto de substantivo quanto de adjetivo, dependendo do contexto, por exemplo: ‘indígenas’, ‘jovens’, ‘doméstica’, ‘locais’, ‘idosos’, ‘brasileiros’. A classificação de tais ocorrências em cada uma dessas classes gramaticais foi realizada por meio da observação das linhas de concordância dos TO e TT e por meio da ferramenta *Concord*, que exhibe cada palavra em contexto. Entretanto, considerando o enfoque dado ao presente estudo, a análise de tal diferença para todas as ocorrências não seria relevante, e será apresentada apenas nos casos dos marcadores culturais que se enquadram neste caso.

Destacamos a seguir as palavras (substantivos, e adjetivos selecionados) mais frequentes após a edição das listas e eliminação das palavras gramaticais e outras fora do nosso escopo de interesse:

Tabela 5 – Trinta maiores ocorrências das listas de palavras dos *releases* em português e inglês, após edição

Palavra (TO)	Frequência	Palavra (TT)	Frequência
anos	378	<i>population</i>	451
população	356	<i>census</i>	361
domicílios	338	<i>persons</i>	341
censo	332	<i>years</i>	255
pessoas	315	<i>areas</i>	230
IBGE	196	<i>income</i>	219
rendimento	183	<i>households</i>	209
mulheres	167	<i>age</i>	198
municípios	165	<i>IBGE</i>	197
região	165	<i>percentage</i>	192
trabalho	165	<i>units</i>	192
indígenas	163	<i>women</i>	183
idade	159	<i>municipalities</i>	173
país	152	<i>region</i>	166
áreas	138	<i>household</i>	145
homens	131	<i>group</i>	144
Norte	130	<i>housing</i>	143
regiões	125	<i>men</i>	139
aglomerados	122	<i>number</i>	121
Rio	109	<i>rural</i>	120
Nordeste	107	<i>Brazil</i>	119
Sul	103	<i>North</i>	118
informações	103	<i>total</i>	116
proporção	103	<i>country</i>	115
nível	102	<i>area</i>	114
Brasil	98	<i>agglomerates</i>	112
percentuais	96	<i>area</i>	111
Sudeste	95	<i>education</i>	109
dados	94	<i>children</i>	107
terras	89	<i>Northeast</i>	107

Destacamos, dentre as trinta maiores ocorrências da tabela acima, em português: ‘aglomerados’, na 19ª posição, que, apesar de ser um adjetivo em contextos diversos, ocorre no subcorpus *releases* apenas como substantivo, e na maioria das vezes (86 casos) acompanhado do adjetivo ‘subnormais’. Tal observação foi feita a partir dos resultados gerados pela ferramenta *Concord*, do programa *WordSmith Tools*. A nomenclatura ‘aglomerado subnormal’ é utilizada pelo IBGE para

designar o que se entende por ‘favelas’, ou ‘comunidades’¹³, e, embora não se possa afirmar ser esse o objetivo da instituição, pode ser percebido como uma forma atenuadora do caráter mais negativo atribuídos às moradias dessa natureza. Há, portanto, um aspecto claramente terminológico, e possivelmente ideológico, na adoção desse vocábulo em especial.

Podemos ver, na tabela acima, possíveis correspondentes entre as maiores ocorrências em português e inglês, entre as quais destacamos: ‘censo’/‘*census*’, ‘população’/‘*population*’ e ‘pessoas’/‘*persons*’. Esses três exemplos podem ser considerados o centro da temática do Censo, uma vez que se referem ao nome da operação, propriamente dita, e à sua “matéria-prima”, ou seja, o povo retratado. Tal predominância temática se reflete na posição desses elementos entre os cinco mais frequentes, ainda que com certa variação em número, tanto no TO quanto no TT.

Os resultados para o texto do questionário encontram-se na tabela a seguir:

¹³ Aglomerado subnormal: Conjunto constituído por um mínimo de 51 unidades habitacionais, ocupando ou tendo ocupado, até período recente, terreno de propriedade alheia (pública ou particular), dispostas, em geral, de forma desordenada e densa, e carentes, em sua maioria, de serviços públicos essenciais. A identificação e a delimitação desses setores especiais são realizadas por ocasião da preparação da base operacional geográfica do Censo Demográfico. Este conceito, utilizado pela primeira vez no Censo Demográfico 1991, possui certo grau de generalização, de forma a abarcar a diversidade de assentamentos irregulares existentes no País, conhecidos como: favela, invasão, grota, baixada, comunidade, vila, ressaca, mocambo, palafita, entre outros. Para o Censo Demográfico 2010, apesar de o conceito de aglomerado subnormal ter permanecido o mesmo de Censos Demográficos anteriores, o IBGE adotou inovações metodológicas e operacionais com o objetivo de atualizar e aprimorar a identificação dos aglomerados. (Glossário do Censo 2010, p.1)

Tabela 6 – Trinta maiores ocorrências das listas de palavras dos questionários em português e inglês, após edição

Palavra (TO)	Frequência (TO)	Palavra (TT)	Frequência (TT)
ano	37	<i>unit</i>	33
domicílio	27	<i>July</i>	26
julho	25	<i>housing</i>	25
trabalho	22	<i>country</i>	20
país	21	<i>municipality</i>	19
município	19	<i>education</i>	17
série	17	<i>year</i>	17
peessoa	16	<i>federation</i>	16
unidade	12	<i>person</i>	15
morador	12	<i>residents</i>	12
federação	12	<i>job</i>	11
dificuldade	12	<i>resident</i>	11
filho	10	<i>work</i>	11
semana	9	<i>birth</i>	9
nome	9	<i>difficulty</i>	9
mês	9	<i>month</i>	9
moradores	8	<i>name</i>	9
idade	8	<i>week</i>	9
rendimento	7	<i>use</i>	8
propriedade	7	<i>household</i>	7
peessoas	7	<i>indigenous</i>	7
horas	7	<i>property</i>	7
hora	7	<i>school</i>	7
filhos	7	<i>adult</i>	6
ensino	7	<i>course</i>	6
curso	7	<i>income</i>	6
uso	6	<i>nursery</i>	6
sexo	6	<i>public</i>	6
responsável	6	<i>sex</i>	6
nascimento	6	<i>youth</i>	6

De maneira análoga ao que se observa nos *releases*, a tabela mostra possíveis correspondentes tradutórios como ‘julho’/’*July*’, ‘país’/’*country*’ e ‘município’/’*municipality*’.

Uma ocorrência da lista merece especial atenção, a palavra ‘ano’, que aparece 37 vezes em inglês e tem seu possível correspondente ‘*year*’ presente apenas 17 vezes no TT. Inicialmente, duas possibilidades se apresentam para explicar tal diferença. A primeira é que nem sempre a idade em inglês agrega a palavra ‘*year*’, uma vez que é possível expressá-la utilizando apenas o número correspondente a idade de alguém, como no caso de “Ela tem 12 anos (de idade)./’*She is 12.*” A segunda possibilidade é que a palavra ‘ano’ tenha um significado tal no texto em português que permita adotar outro correspondente que não ‘*year*’ em inglês.

Uma varredura do questionário em português pela ferramenta *Concord* mais uma vez ajuda a esclarecer a questão, pois mostra que em 13 das ocorrências de 'ano', a palavra remete à série escolar, e não a uma forma de contagem de tempo. Observa-se, ainda, que o possível sinônimo para 'ano', 'série', também aparece na lista de palavras em português, mas seu provável correspondente no mesmo contexto 'grade' não figura na lista dessas primeiras trinta ocorrências.

Para melhor examinar o que ocorre em tais casos, novamente recorremos aos resultados obtidos por meio da ferramenta *Concord* para identificar o contexto de ocorrência de 'série', e concluímos que foi usada com o sentido de 'nível escolar' em todos os casos.

A lista de palavras também evidencia a importância do mês de julho no texto do questionário, sendo 25 ocorrências em português e 26 em inglês para o possível correspondente 'July'. Este mês aparece frequentemente em ambos os idiomas porque marca as datas de corte para diversos quesitos do Censo brasileiro, como podemos ver no exemplo a seguir:

TO: 3.01 - Alguma pessoa que morava com você(s) estava morando em outro país em 31 de julho de 2010?

1 - Sim.

2 - Não.

TT: 3.01 - *Was any person who used to live with you living in another country on July 31, 2010?*

1 - Yes.

2 - No.

Nos subitens a seguir apresentamos resultados mais específicos a respeito da tradução, incluindo exemplos que abarcam nosso interesse principal nesta pesquisa: a identificação e tradução dos marcadores culturais.

4.3 Palavras-chave dos textos originais e traduzidos

A seleção dos marcadores culturais a serem analisados a partir do corpus de estudo, formado por *releases* e questionário, foi inicialmente feita entre as palavras-chave, que resultam do cruzamento entre as listas de palavras e os corpora de referência em português e inglês, geradas por meio da ferramenta *KeyWords*, do *WST*.

As Tabelas 7 e 8 apresentam valores relativos às palavras-chave para os *releases* e o questionário, respectivamente. No primeiro caso, *releases*, foram encontradas 414 palavras-chave no texto em português, e 411 no texto em inglês, ou seja, valores muito próximos. No entanto, quando verificamos a porcentagem que tais palavras-chave representam em relação ao número de formas (*types*), percebe-se que há maior a especificidade no texto em inglês ao idioma geral, uma vez que as palavras-chave em português representam 8,74% do total de formas, e nos TT, 11,78%. Tal observação pode indicar a presença de ocorrências específicas, não tão comuns ao universo da língua inglesa, entre as quais se encontrariam os marcadores culturais.

Tabela 7 – Percentual de palavras-chave dos *releases* em relação à língua geral no TO e no TT

Releases	Nº de formas (<i>types</i>)	Nº de palavras-chave	% de palavras-chave
TO	4.373	414	8,74%
TT	3.488	411	11,78%

Para o questionário, as porcentagens de chavidade são exibidas a seguir:

Tabela 8 – Percentual de palavras-chave dos questionários em relação à língua geral no TO e no TT

Questionário	Nº de formas (<i>types</i>)	Nº de palavras-chave	% de palavras-chave
TO	623	35	5.62%
TT	574	25	4.36%

Observa-se que, ao contrário do que ocorre nos *releases*, a tradução para inglês do questionário apresenta diminuição da porcentagem de palavras-chave em relação ao original, e que os valores são mais próximos, ou seja, há menor variação de representatividade entre as palavras-chave em português e inglês, o que pode

reforçar o argumento relativo às diferenças de natureza entre os textos do tipo *release* e questionário.

4.4 Seleção, significado e classificação dos marcadores culturais

4.4.1 Marcadores culturais presentes nos *releases*

No que se refere à análise dos marcadores propriamente ditos, optamos por selecionar suas ocorrências até o limite máximo de seis por cada domínio cultural investigado: ecológico, ideológico, material e social, seguindo a ordem decrescente de chavicidade, expressa pela coluna *Log_L*. A seguir, apresentamos apenas os marcadores culturais identificados, com o respectivo grau de chavicidade e o domínio cultural a que pertencem:

Tabela 9 – Marcadores culturais identificados entre as palavras-chave dos *releases* em português, em ordem decrescente de chavicidade e por domínio cultural

Marcador	Frequência	Log_L	Domínio cultural
indígenas	163	934,00	social
aglomerados	122	866,55	material/social
regiões	125	430,74	social
municípios	165	405,93	social
região	165	359,37	social
pardos	38	263,97	social
indígena	46	214,54	social
município	64	148,96	social
logradouros	17	101,17	social
logradouro	12	80,25	social
carteira	23	74,20	material
distrito	23	69,81	social
grotas	6	45,82	ecológico
estatutário	7	41,38	social
baixadas	6	40,13	ecológico
pretas	5	38,19	social
distritos	7	34,36	social
mocambos	4	30,55	material
ressacas	4	30,55	ecológico

Embora a lista de palavras-chave nos auxilie na identificação e ordenamento dos marcadores culturais, ela não apresenta todos os marcadores presentes nos *releases* do censo. Por essa razão, para esgotar as possibilidades de ocorrência,

recorremos também à lista de palavras editada, da qual puderam ser extraídos outros marcadores culturais, listados abaixo:

Tabela 10 – Marcadores culturais presentes apenas na lista de palavras dos *releases* em português, em ordem decrescente de frequência

Marcador (TO)	Frequência	Domínio cultural
estados	65	social
estado	36	social
cartório	10	social
favelas	7	material/social
bairros	7	social
bairro	4	social
maloca	4	material/social
oca	4	material/social
palafitas	4	material/social
igarapés	3	ecológico
Yanomami	3	social
candomblé	2	ideológico
malocas	2	material/social
ocas	2	material/social
umbanda	2	ideológico
açudes	1	ecológico
Mundukuru	1	social
Pataxó	1	social
pinguelas	1	material
Sateré Mawé	1	social
Tupinambá	1	social
Xavante	1	social

Ainda em relação à Tabela 10, pode-se observar que alguns elementos foram classificados como pertencentes a dois domínios culturais. Os itens nesta condição são os que designam os tipos de moradias classificadas como exemplos de aglomerados subnormais. Ao mesmo tempo em que descrevem o aspecto físico de uma localidade de condições precárias, também se relacionam a um tipo de morador e perfil social específico em cada caso. Com efeito, segundo tratado por Aubert (2006), e já explicitado em nossa fundamentação teórica, a identificação dos marcadores não é tarefa óbvia, uma vez que eles podem variar segundo o corpus e, ainda, de acordo com as próprias culturas que estão sendo comparadas.

Assim, nossa seleção de marcadores culturais parte da observação dos vocábulos considerados mais particulares da cultura brasileira, em termos de organização administrativa e social, características do relevo, aspectos religiosos e construções específicas dos espaços urbano e rural, itens igualmente presentes nas

realidades investigadas pelo Censo Demográfico. Por questões metodológicas, optamos por desconsiderar todas as ocorrências relativas aos topônimos, como nomes de estados, municípios, etc., que também constituem um grupo culturalmente marcado. Esta decisão baseou-se no fato de que tais vocábulos tendem a estar muito presentes em textos com a temática Censo Demográfico, e incluí-los na listagem a ser analisada inviabilizaria a abordagem de outros possíveis marcadores culturais, dado o limite de seis estabelecido para fins desta pesquisa. Os topônimos encontrados no texto estão listados no Apêndice A.

Em virtude deste cenário, e de forma a justificar as escolhas apresentadas, incluímos, junto aos comentários para cada marcador analisado na seção 4.5, uma breve definição ou contextualização de uso no contexto e textos do Censo Demográfico. As definições para os marcadores listados, mas não analisados, está disponível no Apêndice B.

É possível observar nas Tabelas 9 e 10 a ocorrência de pares do tipo singular/plural, ainda que em posições distantes, os quais consideraremos, para fins de análise, como um único marcador. Assim, havendo sido estabelecido o recorte para a pesquisa de até seis ocorrências por domínio, nossa lista final para análise será composta dos seguintes itens:

- domínio da cultura ecológica: grotas, baixadas, ressacas, igarapés, açudes (5 marcadores);
- domínio da cultura ideológica: candomblé, umbanda (2 marcadores);
- domínio da cultura material: aglomerados, carteira, mocambos, favelas, maloca/malocas, oca/ocas (6 marcadores); e
- domínio da cultura social: indígenas/indígena, regiões/região, municípios/município, pardos, logradouros/logradouro, distritos (6 marcadores).

4.4.2 Marcadores culturais presentes no questionário

A lista de palavras chave do questionário apresenta os seguintes marcadores culturais:

Tabela 11 – Marcadores culturais identificados entre as palavras-chave do questionário em português, em ordem decrescente de chavicidade e por domínio cultural

Marcador (TO)	Frequência	Log_L	Domínio cultural
município	19	173,51	social
supletivo	4	39,06	social

De modo análogo ao que foi feito para os *releases*, também recorreremos à lista de palavras na tentativa de identificar outros marcadores, apresentados abaixo:

Tabela 12 – Marcadores culturais presentes apenas na lista de palavras do questionário em português, em ordem decrescente de frequência

Marcador (TO)	Frequência	Domínio cultural
estado	6	social
indígena	4	social
carteira	2	material
açudes	1	ecológico
cartório	1	social
igarapés	1	ecológico
logradouro	1	social

Mantendo o mesmo parâmetro de seleção anterior, de um limite máximo de seis marcadores por domínio cultural, analisaremos os seguintes:

- domínio da cultura ecológica: açudes, igarapés (2 marcadores)
- domínio da cultura ideológica: não há ocorrências¹⁴
- domínio da cultura material: carteira (1 marcador)
- domínio da cultura social: município, supletivo, estado, indígena, cartório, logradouro (6 marcadores)

Uma breve observação das Tabelas 9, 10, 11 e 12 nos mostra que o domínio da cultura social é predominante entre os marcadores culturais dos *releases* e do questionário, enquanto o domínio da cultura ideológica é o menos observado. Para uma melhor visualização dessa característica, temos a seguir a distribuição relativa dos marcadores culturais em ambos os tipos de texto. Vale a pena ressaltar que, para

¹⁴ Os marcadores do domínio ideológico identificados nos *releases*, 'umbanda' e 'candomblé', não figuram entre as ocorrências do questionário porque a pergunta relativa à religião é aberta, e não apresenta opções ao respondente. Portanto, aparecem apenas nos *releases* porque foram mencionados nas respostas dadas pelos entrevistados acerca deste quesito.

fins deste cálculo, foram considerados os marcadores das Tabelas 9 e 10, no caso dos *releases*, e 11 e 12, no caso do questionário, ou seja, os marcadores culturais provenientes tanto das listas de palavras-chave quanto da lista geral de ocorrências.

Tabela 13 – Distribuição percentual dos marcadores culturais por domínio nos *releases*

Domínio cultural	Porcentagem (%)
Ecológico	11,8%
Ideológico	5,9%
Material	29,4%
Social	52,9%

Tabela 14 – Distribuição percentual dos marcadores culturais por domínio no questionário

Domínio cultural	Porcentagem (%)
Ecológico	12,5%
Ideológico	0,0%
Material	12,5%
Social	75,0%

Os dados acima evidenciam dois aspectos já mencionados acerca dos textos. Em primeiro lugar, é sabido que o Censo Demográfico está associado à vida do povo brasileiro e, deste modo, faz sentido que os marcadores nele presentes tenham predominância do caráter social. Uma segunda observação relevante é que, sendo o texto do questionário mais limitado e específico, a presença dos marcadores do domínio cultural social é ainda mais acentuada.

4.5 Análise das traduções dos marcadores culturais

4.5.1 Análise das traduções dos marcadores dos *releases*

Nesta seção trataremos de aspectos mais específicos das opções tradutórias correspondentes a cada marcador cultural selecionado para análise. Primeiramente abordaremos os marcadores presentes nos *releases*, dispostos na tabela a seguir em ordem alfabética, com as respectivas porcentagens de cada solução tradutória recebida no TT.

Tabela 15 – Marcadores culturais presentes nos *releases* originais, com frequências, e respectivas traduções, com frequências e porcentagens de ocorrência nos *releases* traduzidos para o inglês

Marcador cultural (TO)	Frequência (TO)	Tradução (TT)	Frequência (TO)	Percentual (TT)
açudes	1	dams	1	100,0%
aglomerado	5	agglomerate agglomerates	4 1	80,0% 20,0%
aglomerados	122	agglomerates subnormal agglomerate agglomerations sem tradução*	106 6 4 2 4	87,0% 4,80% 3,28% 1,64% 3,28%
baixadas	6	slums in glens low-lands baixadas (slums in low-lands) slums in low- lands	2 2 1 1	33,0% 33,0% 17,0% 17,0%
candomblé	2	candomblé	2	100%
carteira (assinada/de trabalho assinada)	23	formal contract employment record card formal employment contract formal labor contract	12 6 4 1	52,2% 26,0% 17,5% 4,30%
distritos	7	districts counties	6 1	85,7% 14,3%
favelas	7	favelas favelas (slums) slums	5 1 1	71,4% 14,3% 14,3%
grotas	6	caves grotas (slums in deep valleys) slums in deep valleys	2 2 2	33,3% 33,3% 33,3%
igarapés	3	creeks omissão	2 1	67,0% 33,0%
indígena	46	indian indians indigenous indian-related omissão	28 9 7 1 1	61,0% 19,6% 15,0% 2,2% 2,2%
indígenas	163	indians indian	79 75	48,5% 46,0%

		indigenous omissão	8 1	4,9% 0,6%
logradouro	12	address places street omissão	9 1 1 1	75,0% 8,33% 8,33% 8,33%
logradouros	17	omissão address areas	12 3 2	70,0% 18,0% 12,0%
maloca	4	malocas maloca	2 2	50,0% 50,0%
malocas	2	malocas	2	100,0%
mocambos	4	mocambo (type of shack) type of shack	3 1	75,0% 25,0%
município	64	municipality municipalities municipal omissão	47 10 4 3	73,4% 15,7% 6,3% 4,6%
municípios	165	municipalities municipal sem tradução* municipality omissão	155 5 3 1 1	94,0% 3,0% 1,8% 0,6% 0,6%
pardo	1	brown	1	100,0%
pardos	38	brown browns	32 6	84,2% 15,8%
região	151	region omissão major regions regions	136 8 4 3	90,0% 5,3% 2,7% 2,0%
regiões	96	regions major regions omissão do marcador areas	48 19 25 4	50,0% 19,8% 26,0% 4,2%
grandes regiões	13	major regions regions areas	10 2 1	77,0% 15,4% 7,6%
região metropolitana	14	metropolitan area metropolitan areas	13 1	92,9% 7,1%
regiões metropolitanas	16	metropolitan areas	16	100,0%
ressacas	4	slums in back waters ressacas (slums in back waters)	3 1	75,0% 25,0%
umbanda	2	umbanda	2	100,0%

*Os itens classificados como “sem tradução” apareceram no texto como endereços de Internet, ou seja, foram contabilizados pelo programa *WordSmith Tools* mas não tinham significado no texto.

Apresentamos, a seguir, alguns comentários e exemplos de cada caso.

Açudes

Vocábulo apontado como brasileirismo em uma de suas acepções, segundo a atual versão do *Dicionário Michaelis Online*: “espécie de lago que se forma por represamento”.

Em sua única ocorrência nos *releases*, o vocábulo ‘açudes’ foi adaptado para ‘dams’, tradução que, de certa forma, generaliza o significado original. Conserva-se a ideia de espaço de água doce, mas fica comprometida a comparação que se apresenta no original e diferencia ‘lagos’ de ‘açudes’.

TO: A região liderava na categoria “outra forma de abastecimento”, com 44,6%. Desse contingente, 85,1% vinham de rios, açudes, lagos e igarapés.

TT: *The region was the leader in the category “other type of supply”, with 44.6%. From this total, 85.1% came from rivers, dams and lakes.*

O professor britânico consultado para a leitura de alguns textos não conhecia o conceito de açude. Após alguns esclarecimentos incluindo a busca de imagens no Google, sugeriu como possibilidade de tradução a palavra de origem francesa ‘reservoir’, pois esta seria mais indicada para referir-se à água doce acumulada, proveniente ou não de uma represa.

Aglomerado/aglomerados

Este é o termo utilizado pelo IBGE para designar as chamadas ‘comunidades’ ou ‘favelas’. Neste caso, temos uma questão cultural “duplicada”: em primeiro lugar, a inexistência, em outras culturas, de ‘favelas’ como as conhecemos, e a ausência de termo técnico equivalente para designá-las em um instituto de pesquisa estrangeiro. Em segundo lugar, o fato de que ‘aglomerado’ pode causar estranhamento até mesmo para falantes do português, uma vez que se trata de um termo cunhado pelo próprio IBGE, e não uma palavra de uso comum.

Os resultados apresentados pelo *WST* mostram a combinação do vocábulo ‘aglomerados’ ao adjetivo ‘subnormais’, em mais de 70% dos casos. Nos outros quase 30%, houve a supressão do adjetivo, em geral por menção anterior no mesmo período.

Assim, tais observações apontam para o uso do colocado ‘aglomerado subnormal’ como referente às comunidades e/ou favelas ao redor do Brasil.

Uma das opções de tradução encontradas para ‘aglomerado’ foi a tradução literal ‘*agglomerate*’, que mantém a neutralidade no TT em relação ao TO:

TO: O conceito de aglomerado subnormal foi utilizado pela primeira vez no Censo Demográfico de 1991.

TT: *The concept of subnormal agglomerate was used for the first time in the 1991 Population Census*

Observou-se também a transposição para a forma plural ‘*agglomerates*’, que confere ao marcador um sentido mais genérico, conforme exemplificado seguir:

TO: Em Belém, uma das características dominantes é a grande extensão das áreas de aglomerado subnormal.

TT: *A relevant feature in Belém is the large extension of the areas of subnormal agglomerates.*

No caso de ‘aglomerados’, no plural, a tradução literal ‘*agglomerates*’ foi predominante, sendo adotada em mais de 90% das ocorrências:

TO: Em outubro, serão publicados os resultados definitivos do Universo, inclusive por setor censitário, além de informações sobre aglomerados subnormais e indicadores sociais municipais básicos.

TT: *October will bring the final Universe results, including information by enumeration area, besides data about subnormal agglomerates and basic municipal social indicators.*

A transposição para ‘*agglomerate*’ também foi observada, conforme o exemplo abaixo:

TO: A publicação completa será disponibilizada na página do IBGE, além de tabelas com as informações para cada um dos aglomerados e arquivos com seus limites em formato shape file.

TT: *The full publication will be available on the web page of IBGE, including tables containing information for each agglomerate and files with their limits in shape file.*

Houve, ainda, reduzidas porcentagens de transposição para o substantivo ‘*agglomerations*’, para o adjetivo ‘*subnormal*’ para remeter ao significado de ‘aglomerados’, e de omissão do marcador cultural, conforme os três exemplos abaixo:

TO: Da mesma forma, as áreas rurais podem ser classificadas como aglomerados rurais de extensão urbana, povoados, núcleos ou outros aglomerados, todos eles também definidos por legislação municipal.

TT: *Likewise, the rural areas can be classified as rural agglomerations of urban extension, villages, nucleus or other agglomerations; all of them also defined by municipal legislation.*

TO: Vinte regiões metropolitanas concentravam 88,6% desses domicílios, e quase metade (49,8%) dos domicílios de aglomerados estavam na Região Sudeste.

TT: *Twenty metropolitan areas concentrated 88.6% of these subnormal households, and almost half of them (49.8%) were located in the Southeast Region.*

TO: Na Região Metropolitana de Belém 87,0% dos domicílios em aglomerados subnormais tinham acesso por ruas

TT: *87.0% of the housing units in the metropolitan area of Belém were located in areas predominantly accessed through streets.*

As traduções ‘*agglomerate*’ e ‘*agglomerates*’ não foram consideradas muito pertinentes pelo consultor Donal Hunter, pois, apesar de existir uma asseção relacionada à geografia nas definições de tais palavras, não fica muito claro ao que se referem tais vocábulos. Como sugestão foi mencionada a palavra ‘*conurbation*’.

Baixadas

As baixadas, como o próprio nome anuncia, são áreas baixas – e geralmente alagadiças – onde parte da população em vulnerabilidade social constrói suas moradias; estão incluídas no conjunto definido pelo IBGE como aglomerados subnormais.

É importante observar que, a despeito do menor número de ocorrências em relação aos demais, os marcadores que descrevem aglomerados subnormais em geral apresentam significativa variação de escolhas tradutórias. Com relação ao vocábulo ‘baixadas’, foram adotadas quatro opções, sendo que em três delas a expressão ‘*low-lands*’ é contemplada de alguma forma, seja isoladamente ou em associação a outro vocábulo, com exceção do caso de ‘*slums in glens*’. O exemplo abaixo traz um caso em que observamos a adoção de duas modalidades simultaneamente para traduzir o referido marcador: empréstimo, seguido de modulação:

TO: Em 2010, o país possuía 6.329 aglomerados subnormais (assentamentos irregulares conhecidos como favelas, invasões, grotas, baixadas, comunidades [...])

TT: *In 2010 Brazil had 6,329 subnormal agglomerates in 2010 – irregular settlements known as favelas (slums), invasões (invaded properties), grotas (slums in deep valleys), baixadas (slums in low-lands), comunidades (poor communities) [...]*

No exemplo a seguir observamos somente a modulação para ‘*low-lands*’:

TO: Os aglomerados subnormais frequentemente ocupam áreas menos propícias à urbanização, como encostas íngremes no Rio de Janeiro, áreas de praia em Fortaleza, vales profundos em Maceió (localmente conhecidos como grotas), baixadas permanentemente inundadas em Macapá, manguezais em Cubatão, igarapés e encostas em Manaus.

TT: *The subnormal agglomerates are often settled in areas less appropriate to urbanization, like steep hillsides in Rio de Janeiro, beach areas in Fortaleza, deep valleys in Maceió – locally known as grotas –, permanently flooded low-lands in Macapá, swamps in Cubatão, creeks and hillsides in Manaus.*

A opção ‘*low-lands*’ foi considerada adequada pelo consultor. Por outro lado, ‘*slow in glens*’ não lhe pareceu uma opção adequada pois ‘*glens*’ refere-se a um vale típico da Escócia.

Candomblé / Umbanda

Estes dois marcadores têm poucas ocorrências e aparecem no texto juntos, por isso são apresentados aqui da mesma maneira, e ambos referem-se às religiões comumente chamadas “culto afro-brasileiros”.

O candomblé é originário da Nigéria e Benin, e foi trazido pelos negros africanos que foram traficados e escravizados em nosso país. Atualmente, a religião engloba também seitas derivadas candomblé original.

A umbanda é uma religião essencialmente brasileira, de natureza sincrética, que mistura elementos de religiões africanas e cristãs.

O marcador ‘candomblé’ tem apenas duas ocorrências e, em ambos os casos, apresenta-se em par com o marcador ‘umbanda’, em referência às religiões de matriz africanas contempladas pelo Censo 2010. Optou-se, em 100% das traduções, pela manutenção dos vocábulos originais, ou seja, pelo empréstimo, não havendo a inclusão de notas explicativas ou qualquer outro esclarecimento acerca do significado dos termos além do próprio contexto de ocorrência:

TO: Em 2000 eram quase 12,5 milhões (7,3%), ultrapassando os 15 milhões em 2010 (8,0%). Os adeptos da umbanda e do candomblé mantiveram-se em 0,3% em 2010.

TT: *In 2000 this group was formed by 12.5 million persons (7.3%), and surpassed the 15 million of 2010 (8.0%). The followers of umbanda and candomblé remained at 0.3% in 2010.*

TO: A maior representatividade de pretos foi verificada na umbanda e candomblé (21,1%).

TT: *Blacks were the majority in umbanda and candomblé (21.1%).*

Segundo o consultor, o candomblé é reconhecido como algo relacionado à cultura e religiosidade africanas, então não seria difícil compreender seu sentido texto. Curiosamente, para ele, a palavra ‘umbanda’ não evocou nenhuma ideia, uma vez que ele desconhecia completamente do que se tratava. No entanto, pôde identificar a palavra como sendo parte do universo religioso brasileiro, dado o seu contexto.

Carteira

Nas ocorrências encontradas nos textos do Censo 2010, o vocábulo ‘carteira’ refere-se à carteira de trabalho, documento necessário para a oficialização da atividade laboral no setor privado brasileiro. A investigação acerca da chamada “assinatura” da carteira de trabalho, ou seja, do estabelecimento de vínculo empregatício formal, também fornece dados acerca das relações de trabalho no Brasil.

No caso deste marcador, a maioria das ocorrências em português subdivide-se entre duas expressões: ‘carteira assinada’ e ‘carteira de trabalho assinada’. As traduções variaram em quatro opções, sendo a mais adotada, em termos percentuais, a modulação para ‘*formal contract*’ (52% dos casos):

TO: Entre os empregados, os com carteira de trabalho assinada foram 67,7%, no contingente masculino, e 59,3%, no feminino.

TT: *Among the employees, those with a formal contract were 67.7%, among men, and 59.3%, among women.*

A segunda opção mais observada, a modulação ‘*employment record card*’, com 26%, refere-se ao documento propriamente dito:

TO: Os percentuais de trabalhadores com deficiência por conta própria (27,4%), sem carteira (22,5%) [...]

TT: *The percentage of disabled workers who were self-employed (27.4%), without a signed employment record card (22.5%) [...]*

A terceira e a quarta opções mais frequentes (‘*formal employment contract*’ e ‘*formal labor contract*’) são equivalentes entre si e deixam explícita a referência ao contexto de trabalho, não ao documento, e também adotam o vocábulo ‘*formal*’ para referir-se à situação regularizada dos trabalhadores que possuem o registro em carteira:

TO: No total da população ocupada do País, os empregados com carteira de trabalho assinada representavam 44,5% [...]

TT: *Considering the total employed population in the country, workers with a formal employment contract represented 44.5% [...]*

TO: De 2000 para 2010, no subgrupo dos trabalhadores domésticos, a categoria dos com carteira de trabalho assinada cresceu de 29,8% para 33,8%.

TT: *From 2000 to 2010, in the domestic workers subgroup, the category with a formal labor contract grew from 29.8% to 33.8%.*

Todas as traduções apresentadas foram consideradas pertinentes pelo consultor. No entanto, o mesmo também sugeriu a expressão “*fully employed*” para se referir aos trabalhadores que possuem vínculo empregatício formal.

Distritos

‘Distritos’ foi selecionado como marcador cultural por referir-se a uma divisão administrativa e territorial que pode abranger diversos bairros. Como outras classificações do tipo, não encontra equivalente pleno em outras culturas/línguas.

A tradução de distritos foi feita, na maioria dos casos presentes nos *releases*, com a tradução literal ‘*districts*’, segundo nos mostra o exemplo abaixo:

TO: Nos 40 municípios brasileiros que possuem mais de 500 mil habitantes, de modo geral 5% dos domicílios responderão ao questionário da Amostra, exceto em alguns distritos que terão frações amostrais maiores.

TT: *In the 40 Brazilian municipalities which have more than 500 inhabitants, in general, 5% of the households will answer the Sample questionnaire, except some districts with bigger sampling fractions.*

É interessante observar, no entanto, que um mesmo trecho em português foi incluído em dois *releases* diferentes, e que as opções tradutórias se modificaram, sendo uma adaptação para ‘*counties*’, na primeira, e a tradução literal, ‘*districts*’, na segunda:

TO: O Censo Demográfico é a única pesquisa domiciliar a levantar informações sobre todos os 5.565 municípios brasileiros e, para alguns quesitos, trará dados até mesmo para bairros e distritos.

TT: *The Demographic Census is the only household survey to assess information about all the 5,565 Brazilian municipalities, bringing, for some aspects, specific data about neighborhoods and counties.*

TO: O Censo Demográfico é a única pesquisa domiciliar a levantar informações sobre todos os 5.565 municípios brasileiros e, para alguns quesitos, trará dados até mesmo para bairros e distritos.

TT: *The Demographic Census is the only household survey to assess information about all the 5,565 Brazilian municipalities and, for some criteria, it will even bring data about neighborhoods and districts.*

Não houve questionamentos pelo consultor quanto ao uso de 'districts'.

Favelas

O nome 'favela' é proveniente de um tipo de planta encontrada no Brasil, e designa um conjunto de habitações populares amontoadas, em locais de ocupação irregular, em geral, com estrutura precária; as favelas estão incluídas no conjunto definido pelo IBGE como aglomerados subnormais.

Como tradução do marcador cultural 'favelas', foram encontradas a palavra em inglês 'slums', adaptação que aproxima-se do significado das nossas chamadas comunidades, e o empréstimo 'favela', que atualmente já faz parte de dicionários de língua inglesa, por exemplo, o *Webster Online* ("a settlement of jerry-built shacks lying on the outskirts of a Brazilian city") e o *Oxford Online* ("A Brazilian shack or shanty town; a slum."). Abaixo vemos dois exemplos:

TO: Dados do Questionário da Amostra do Censo 2010 evidenciam desigualdades entre a população que residia em aglomerados subnormais (assentamentos irregulares conhecidos como favelas, invasões, grotas, baixadas, comunidades, vilas, ressacas, mocambos e palafitas, entre outros)

TT: *Data from the Long Form Questionnaire of the 2010 Census show inequalities between the population that lived in subnormal agglomerates (irregular settlements known as slums, invaded properties, slums in deep valleys, slums in low-lands, poor*

communities, slums in villages, slums in backwaters, types of shacks and stilt houses, among others).

TO: O Instituto lançou em 1953 o volume “As favelas do Distrito Federal e o Censo Demográfico de 1950”, quando foi apurado que 7,2% da população do Distrito Federal (169.305 pessoas) eram moradores de favelas.

TT: *IBGE released the publication 'The Favelas of the Federal District and the 1950 Population Census in 1953 ', when it was detected that 7.2% of the population of the Federal District (169,305 inhabitants) lived in favelas.*

Embora já haja registros de ‘favelas’ em dicionários da língua inglesa e referências a esse tipo de localidade em veículos de mídia que abordem o Brasil, o professor Donald Hunter apontou ‘*shanty towns*’ como uma possibilidade mais conhecida aos falantes do inglês em geral.

Grotas

Refere-se a um tipo de cavidade em uma encosta ou na ribanceira de um rio, aberta pela água de chuva ou enchentes, e compõe o conjunto definido pelo IBGE como aglomerados subnormais.

Para o marcador ‘grotas’ observam-se quatro diferentes ocorrências tradutórias, que variam desde a manutenção do termo em português, sem qualquer nota adicional, até uma transposição em três vocábulos que parecem não esclarecer o significado, conforme ilustrado abaixo:

TO: Os aglomerados subnormais frequentemente ocupam áreas menos propícias à urbanização, como encostas íngremes no Rio de Janeiro, áreas de praia em Fortaleza, vales profundos em Maceió (localmente conhecidos como grotas), baixadas permanentemente inundadas em Macapá, manguezais em Cubatão, igarapés e encostas em Manaus.

TT: *The subnormal agglomerates are often settled in areas less appropriate to urbanization, like steep hillsides in Rio de Janeiro, beach areas in Fortaleza, deep valleys in Maceió – locally known as grotas –, permanently flooded low-lands in Macapá, swamps in Cubatão, creeks and hillsides in Manaus.*

TO: Dados do Questionário da Amostra do Censo 2010 evidenciam desigualdades entre a população que residia em aglomerados subnormais (assentamentos irregulares conhecidos como favelas, invasões, grotas, baixadas, comunidades, vilas, ressacas, mocambos e palafitas, entre outros)

TT: *Data from the Long Form Questionnaire of the 2010 Census show inequalities between the population that lived in subnormal agglomerates (irregular settlements known as slums, invaded properties, slums in deep valleys, slums in low-lands, poor communities, slums in villages, slums in backwaters, types of shacks and stilt houses, among others)*

Vale a pena observar que, no segundo exemplo acima, os tradutores optaram por partir do vocábulo ‘*slum*’ para definir os demais locais listados, adaptando cada tradução de acordo com o contexto geográfico particular, e, de certa forma, gerando uma neutralização da riqueza de significados trazida pelo texto original.

Não foi apontada pelo consultor uma outra opção para a tradução de ‘grotas’, e uma das traduções presentes no texto, ‘*cave*’, foi considerada bastante adequada.

Igarapés

Adotado da língua tupi, “caminho da canoa”, designa um tipo de canal estreito, suficiente para a travessia de pequenas embarcações e presente principalmente na região amazônica do Brasil

Nas ocorrências de ‘igarapé’ houve um exemplo de omissão do marcador, e a modulação ‘*creeks*’, nos demais casos, conforme exemplificado:

TO: A região liderava na categoria “outra forma de abastecimento”, com 44,6%. Desse contingente, 85,1% vinham de rios, açudes, lagos e igarapés.

TT: *The region was the leader in the category “other type of supply”, with 44.6%. From this total, 85.1% came from rivers, dams and lakes.*

TO: Os aglomerados subnormais frequentemente ocupavam áreas menos propícias à urbanização, como encostas íngremes no Rio de Janeiro, áreas de praia em Fortaleza, vales profundos em Maceió (localmente conhecidos como grotas), baixadas

permanentemente inundadas em Macapá, manguezais em Cubatão, igarapés e encostas em Manaus.

TT: *The subnormal agglomerates were often settled in areas less appropriate to urbanization, like steep hillsides in Rio de Janeiro, beach areas in Fortaleza, deep valleys in Maceió – locally known as grotas -, permanently flooded low-lands in Macapá, swamps in Cubatão, creeks and hillsides in Manaus.*

Neste caso, não houve sugestão do consultor nativo, pois o mesmo não compreendia o sentido pleno de igarapés, e mediante esclarecimento acerca do termos, considerou suficiente a opção '*creeks*'.

Indígena/indígenas

No Brasil, 'indígena(s)' trata de um conjunto de povos nativos que compartilham algumas características, mas cujas tribos também diferenciam-se em outros aspectos, como a língua, costumes, etc. Embora haja populações indígenas em alguns países falantes da língua inglesa, não necessariamente trata-se do mesmo tipo de povo.

Foram registradas várias diferentes opções tradutórias tanto para a forma singular quanto a plural, sendo quatro delas comuns a ambas, com diferenças apenas na porcentagem de ocorrência: '*indigenous*', '*indian*', '*indians*', e omissão do marcador. A escolha mais particular se deu na transposição para '*indian-related*', como tradução de 2.2% das ocorrências de 'indígena'.

Outra diferença que merece menção refere-se ao uso de 'indígena/indígenas' como adjetivo. No singular, a tradução literal '*Indian*' foi predominante; no plural, entretanto, optou-se por '*indigenous*', resultando nos pares mais frequentes 'indígena/*Indian*' e 'indígenas/*indigenous*', conforme ilustrado nos exemplos abaixo:

TO: A população indígena estava concentrada (60,8%) nas áreas rurais, enquanto 15,6% do total da população brasileira vivia nessas áreas.

TT: *The Indian population was concentrated (60.8%) in rural areas, whereas 15.6% of the total Brazilian population lived in these areas.*

TO: Censo investigará línguas indígenas pela primeira vez

TT: *2010 Census will investigate indigenous languages for the first time*

No caso do uso como substantivo, somente observada no caso do plural ‘indígenas’, predominou a tradução ‘Indians’:

TO: Pela primeira vez, o Censo investigará a língua falada pelos indígenas residentes no território nacional.

TT: *For the first time, the Census will investigate the language spoken by Indians living in the national territory*

Segundo observado pelo professor Donald Hunter, a opção mais adequada para o par indígena/indígenas seria o adjetivos ‘*indigenous*’, adaptado a cada situação. Foi mencionada como argumento a preocupação com uma tradução mais politicamente correta, uma vez que o uso de ‘Indians’ pode causar confusão com o povo nascido na Índia e o vocábulo ‘*native*’ para designar apenas essas pessoas pode, segundo ele, soar como pejorativo uma vez que todos os nascidos no país seriam nativos do mesmo modo.

Logradouro/logradouros

Definido no *Dicionário Michaelis Online* (2018) como “lugar livre destinado à circulação pública de pedestres e veículos, tal como ruas, avenidas, praças, viadutos”. O vocábulo logradouro está diretamente relacionado às divisões geográficas dos espaços segundo o entendimento da administração pública do Brasil. Assim como regiões e municípios, surge em um contexto espacial que não encontra definição exata em inglês, uma vez que a forma de organização e classificação do espaço urbano pode não encontrar similar em países estrangeiros.

Foram observadas diferenças significativas nesse par, ou seja, entre as formas singular e plural. Para ‘logradouro’, a tradução ‘*address*’ foi a mais adotada, tendo havido apenas uma omissão, conforme ilustrado abaixo:

TO: A pavimentação das vias (81,7%), meio fio/guia (77,0%), calçadas (69,0%), arborização (68,0%) e identificação do logradouro (60,5%) também apresentaram percentuais superiores a 50% [...]

TT: *Road paving (81.7%), curbs (77.0%), sidewalks (69.0%), forestation (68.0%) and address identification (60.5%) also presented indexes above 50% [...]*

TO: O lixo de todos os domicílios era jogado em terreno baldio ou logradouro em seis terras: Areões (MT), Zo'E (PA), Aripuanã (MT), Badjonkore (PA), Riozinho do Alto Envira (AC) e Mundo Verde/Cachoeirinha (MG).

TT: *The garbage from all housing units was thrown on vacant lots in six areas: Areões (MT), Zo'E (PA), Aripuanã (MT), Badjonkore (PA), Riozinho do Alto Envira (AC) and Mundo Verde/Cachoeirinha (MG).*

No caso do plural, ao contrário, o vocábulo foi omitido em 70,0% dos casos, enquanto as opções 'address' e 'areas' foram utilizadas em apenas 18,0% e 12,0% das ocorrências, respectivamente:

TO: A pesquisa verificou, também, a existência de esgoto a céu aberto (11,0%) e depósitos de lixo nos logradouros (5,0%) do entorno dos domicílios.

TT: *The survey also registered the existence of open sewage (11.0%) and accumulation of garbage (5.0%) in the household surroundings.*

TO: Observou-se que os municípios com mais de um milhão de habitantes apresentaram percentuais mais altos de domicílios localizados em áreas com iluminação pública (97,1%), pavimentação das vias públicas (92,8%), meio fio/guia (85,8%), calçada (82,9%) identificação de logradouros (79,9%) e 8,6% de rampas para cadeirantes.

TT: *Municipalities with more than one million inhabitants presented the highest percentages of households located in areas with public electricity (97.1%), road paving (92.8%), curb (85.8%), sidewalk (82.9%), address identification (79.9%) and 8.6% of wheelchair ramps.*

TO: O percentual de domicílios com depósito/acúmulo de lixo nos logradouros em seu entorno aumentava proporcionalmente em relação ao tamanho do município [..]

TT: *The percentage of household with garbage accumulation in the surrounding areas rose proportionally in relation to the size of the municipality [...]*

Quanto às traduções, o consultor nativo não sugeriu opções para logradouro, embora tenha compreendido, após esclarecimentos, o sentido do vocábulo nos textos do Censo.

Malocas

As malocas são um tipo de cabana comunitária usada como moradia para índios, em especial da região amazônica do Brasil. Esse tipo de habitação pode variar de tribo para tribo.

No caso de ‘malocas’ houve o empréstimo do vocábulo em inglês. Sendo bastante específico da cultura brasileira, ‘maloca’ é atualmente aceito em textos em inglês, e sua definição pode ser encontrada em versões atualizadas de dicionários, em especial os de meio online, como, por exemplo, o Webster Online, que apresenta o verbete como “a large communal dwelling of some South American Indian peoples (as in Brazil); *also*: the group inhabiting such a dwelling”. Abaixo apresentamos exemplos retirados dos textos em português e inglês.

TO: Somente 12,6% dos domicílios eram do tipo “oca ou maloca”, enquanto que, no restante, predominava o tipo “casa”. Mesmo nas terras indígenas, ocas e malocas não eram muito comuns: em apenas 2,9% das terras, todos os domicílios eram desse tipo e, em 58,7% das terras, elas não foram observadas.

TT: *Huts or malocas represented only 12.6% of the housing units, whereas regular houses prevailed in the others. Malocas and huts were not very common, even in areas of Indian land: in only 2.9% of the land, all the households followed this pattern and in 58.7% of the areas, they were not seen.*

Mocambos

A palavra tem origem nos refúgios de escravos fugidos. Refere-se a habitações construídas com materiais improvisados, sempre muito pobre e rústica; está incluída no conjunto definido pelo IBGE como aglomerados subnormais.

Para este marcador foi mantida a palavra original seguida da breve definição ‘*type of shack*’ em todos os casos, porém em 25% deles, o original ‘mocambos’ também estava presente, conforme exemplo:

TO: O IBGE adotou inovações em 2010 para atualizar e aprimorar a identificação dos aglomerados subnormais (assentamentos irregulares conhecidos como favelas, invasões, grotas, baixadas, comunidades, vilas, ressacas, mocambos, palafitas, entre outros).

TT: *IBGE adopted some innovations in 2010 to improve the identification of subnormal agglomerates – irregular settlements such as favelas, invaded properties, caves, slums in glens, poor communities, shantytowns, slums in backwaters, mocambos (type of shack), stilt houses among others.*

Para o professor Hunter, ‘shack’ ou ‘type of shacks são escolhas adequadas pois remetem à rusticidade/precariedade deste tipo de moradia.

Município

Uma vez que a divisão geográfica apresenta diferenças nos vários países, o conceito de ‘município’, como o entendemos, passa a ser um aspecto particular da nossa organização administrativa. Segundo o dicionário *online Michaelis*, uma possível definição para município é uma “circunscrição territorial administrada nos seus próprios interesses por um prefeito, à frente da Prefeitura, que executa as leis emanadas do corpo de vereadores eleitos pelo povo, que compõem a Câmara Municipal.” Compreende-se que não há uma universalidade neste conceito, e que sua tradução para a língua inglesa, quando feita, tende a não abarcar a totalidade de possibilidades nos diversos países falantes do inglês.

Nos exemplos referentes ao Censo 2010, a tradução literal ‘*municipality*’ foi predominante, mas houve pequena opção pela tradução ‘*municipalities*’, no plural, como forma de generalização do caso específico:

TO: A escolha foi aleatória, levando em conta a unidade da federação, o tamanho da população do município e a situação urbana ou rural.

TT: *The choice was random, considering the Federation Unit of location and the population size of the municipality, as well as its urban or rural status.*

TO: Também há de se ressaltar que os aglomerados subnormais podem ter nomes iguais aos de bairros (divisões legais do município), sem que as áreas necessariamente coincidam.

TT: *It is important to mention that a subnormal agglomerate may have the same name of a district – legal division of municipalities – whether or not they match.*

Foram observadas mínimas ocorrências da transposição para ‘municipal’ e de omissão do marcador:

TO: Além da densidade demográfica e da contagem da população – detalhada até o nível de setor censitário e com informações por sexo, grupos de idade e situação de domicílio –, também já foram divulgados resultados preliminares do Universo até o nível de município [...]

TT: *Besides demographic density and population count – detailed at enumeration area level and consisting of information by sex, age groups and household status – there have also been preliminary Universe results at municipal level [...]*

TO: Instaladas a partir de julho de 2009, em 5.564 municípios (exceto no município de Macuco – RJ), tiveram a participação, na segunda reunião, de mais de 60 mil membros cadastrados.

TT: *Starting their activities in July, in 5,564 municipalities (except in Macuco/RJ), the Commissions had, in the second meeting, more than 60 thousand registered members.*

Municípios

Para este marcador foram encontradas duas traduções nos textos analisados: ‘municipalities’ e ‘municipal’. Além desses exemplos, de tradução literal e transposição, observamos omissões da palavra no texto em inglês. O cotejo dos textos em paralelo evidenciou tais ocorrências, exemplificadas abaixo, nesta mesma ordem:

TO: Nos municípios com até 2.500 pessoas (260 municípios), o questionário da Amostra será aplicado em 50% das casas e no restante será respondido o questionário Básico.

TT: *In municipalities with up to 2,500 persons (260 municipalities), the Sample questionnaire will be administered in 50% of the households; the other 50% will answer the Basic questionnaire.*

TO: O Censo também fornece as referências para as projeções populacionais, com base nas quais o Tribunal de Contas da União define as cotas do Fundo Nacional de

Participação dos Municípios, e a partir das quais é definido o número de deputados e vereadores.

TT: *It also provides references for national population prospects, based on which the national Court of Audit defines the shares for the Municipal Participation Fund, and eventually determines the number of deputies and councilmen.*

TO: Representantes dos poderes Executivo, Legislativo e Judiciário dos estados e municípios, além de instituições da sociedade civil, estão acompanhando e auxiliando as atividades do Censo 2010, nas Comissões Censitárias Estaduais (CCE) e nas Comissões Municipais de Geografia e Estatística (CMGEs).

TT: *Representatives of the Executive, Legislative and Judiciary Powers and also institutions belonging to the civil society have been monitoring and helping in the 2010 Census activities in State Census Commissions (CCE) and in the Municipal Commissions of Geography and Statistics (CMGEs).*

De forma geral, a tradução do par município/municípios foi uma das que recebeu mais críticas negativas do consultor. Segundo o britânico, embora o vocábulo ‘municipality’ conste de dicionários em inglês, seu significado não seria de conhecimento do leitor comum. Por ser uma palavra de origem latina, pode levar à ideia de um tipo antigo de organização político-administrativa. Havendo sido discutida com ele a relativa dificuldade de harmonização deste tipo de vocabulário, uma vez que há variação também entre os países falantes de língua inglesa, foi sugerida a opção ‘province’.

Pardo/pardos

As questões de cor e raça costumam causar controvérsia em contextos diversos, e não é diferente no que se refere à classificação do IBGE. Para o Instituto, pardo(s) significa pessoas com uma mistura de cores de pele, resultante da miscigenação cafuza, cabocla ou mulata. A classificação de um indivíduo como pardo na pesquisa do Censo baseia-se na autodeclaração, ou seja, é a própria pessoa que, ao responder o questionário, enquadra-se neste grupo. Por ser uma definição que agrega, ao mesmo tempo, a realidade de cor e raça vivenciada no Brasil e também a

conceituação particular da instituição, ‘pardos’ é considerado um marcador cultural para fins desta pesquisa.

Neste caso a escolha tradutória predominante foi ‘*brown*’, como adjetivo, palavra da língua inglesa que significa, segundo o dicionário *Merriam-Webster Online*: “*of the color brown; especially: of dark or tanned complexion*”. Uma vez que ‘pardo’ abarca boa parte da diversidade de cor observada no país, acreditamos que a opção em inglês tenha sido feita para manter o mesmo aspecto genérico do vocábulo original:

TO: Além disso, o percentual de analfabetos entre pretos (14,4%) e pardos (13,0%) era, em 2010, quase o triplo dos brancos (5,9%).

TT: In addition, in 2010 the percentage of illiterate among black (14.4%) and brown people (13.0%) was almost three times as much as that of the white people (5.9%).

Também observou-se a adoção de ‘browns’ como substantivo:

TO: Os percentuais eram de 25,5% dos brancos, 20,3% dos pretos e 21,6% dos pardos. Na faixa de 15 a 17 anos os valores foram bem maiores, sendo de 85,4% para os brancos, 81,1% para pretos e 81,9% para pardos.

TT: The percentages were of 25.5% of whites, 20.3% of blacks and 21.6% of browns. In the group aged 15 to 17, the values were even greater, 85.4% for the white, 81.1% for the black and 81.9% for the brown.

A opção ‘*brown*’ foi considerada pelo consultor bastante adequada para tentar refletir a parcela da população de cor intermediária entre brancos e negros.

Região/regiões

No contexto brasileiro, o conceito de regiões deixa de ser apenas “qualquer extensão territorial” (segundo o *Dicionário Michaelis Online*), uma vez que designa também a subdivisão dos estados em cinco grupos, segundo sua posição geográfica, também chamadas, pelo IBGE, de “Grandes Regiões”. É importante ressaltar que os textos do Censo também apresentam ocorrências de ‘regiões metropolitanas’, que se

diferenciam do tipo de região descrito anteriormente, e que são recortes instituídos por lei estadual de modo a integrar o planejamento e execução de funções públicas. A definição das chamadas RM é de competência dos estados nos termos do Artigo 25, Parágrafo 3º da Constituição Federal. Segundo dados obtidos do Sistema IBGE de Recuperação automática (SIDRA), acessível por meio do website do Instituto (<https://sidra.ibge.gov.br/territorio#/N7>), o Brasil conta atualmente com 73 regiões metropolitanas. Assim sendo, nossas análises serão realizadas para região/regiões/regiões metropolitanas.

Conforme ilustrado na Tabela 15, cada uma das formas, singular e plural, teve quatro opções de tradução e também houve omissão do marcador em ambos os casos. As soluções tradutórias comuns foram ‘*Major Regions*’ e ‘*regions*’. Uma diferença foi o uso de ‘*areas*’, de forma a generalizar o significado, no plural. Seguem alguns exemplos:

TO: Entre as regiões, o maior contingente ficava na região Norte, 342,8 mil indígenas e o menor no Sul, 78,8 mil. Um total de 517,4 mil (57,7% do total nacional) residiam em terras indígenas, dos quais 251,9 mil (48,7%) estavam na região Norte.

TT: *Considering the Major Regions, the biggest number of Indians was found in the North, 342.8 thousand, and the smallest, in the South, 78.8 thousand. A total 517.4 thousand persons (57.7% of the national total) lived in areas of Indian land, and 251.9 thousand of them (48.7%), in the North Region.*

TO: Para as demais regiões, em 2010, os percentuais eram de 49,4% no Norte, 54,8% no Sudeste e 57,9% no Centro-Oeste.

TT: *The percentages for the other Major Regions in 2010 were: 49.4% in the North, 54.8% in the Southeast and 57.9% in the Central West.*

Região metropolitana/regiões metropolitanas

A tradução ‘*metropolitan area(s)*’ foi adotada como padrão em todas as ocorrências deste marcador. Não há nenhuma referência explicativa adicional acerca do significado da expressão, sendo seu entendimento depreendido apenas do contexto:

TO: As regiões metropolitanas concentraram a maioria dos domicílios em aglomerados subnormais do país, com destaque para a região metropolitana de São Paulo (RMSP) com 596 479 domicílios em setores de aglomerados subnormais.

TT: *The metropolitan areas concentrated the majority of the housing units in subnormal agglomerates in Brazil. The highlight was the metropolitan area of São Paulo (RMSP), which comprised 596,479 housing units in enumeration areas within subnormal agglomerates.*

As opções tradutórias dadas a ‘região/regiões/regiões metropolitanas’ foram consideradas adequadas na visão do leitor nativo consultado.

Ressacas

As chamadas áreas de ressaca comportam-se como reservatórios naturais de água com influência da chuva e maré, onde parte da população em vulnerabilidade social constrói suas moradias, sendo particularmente denominada por tal nome no estado do Amapá; está incluída no conjunto definido pelo IBGE como aglomerados subnormais

A moradia do tipo ‘ressaca’ recebeu a mesma tradução em todos os casos, conforme mostrado no exemplo abaixo:

TO: Dados do Questionário da Amostra do Censo 2010 evidenciam desigualdades entre a população que residia em aglomerados subnormais (assentamentos irregulares conhecidos como favelas, invasões, grotas, baixadas, comunidades, vilas, ressacas, mocambos e palafitas, entre outros) e a que morava nas demais regiões dos municípios, diferenças estas que se evidenciavam também entre as cinco grandes regiões do país.

TT: *Data from the Long Form Questionnaire of the 2010 Census show inequalities between the population that lived in subnormal agglomerates (irregular settlements known as slums, invaded properties, slums in deep valleys, slums in low-lands, poor communities, slums in villages, slums in backwaters, types of shacks and stilt houses, among others) and those living in other areas of the municipalities.*

De maneira similar ao que foi observado em alguns outros marcadores que designam aglomerados subnormais, ‘ressacas’ recebeu a mesma tradução em 100% dos casos: ‘*slums in backwaters*’. Novamente, como em ‘baixadas’, utilizou-se o como base o vocábulo ‘*slum*’, acrescido de um termo explicativo para tentar expressar o significado.

Segundo o professor Donald Hunter, ‘*slums in backwaters*’ é uma escolha satisfatória e, de alguma forma, capaz de dar ao nativo a noção do tipo de espaço físico tratado.

4.5.2 Análise das traduções dos marcadores do questionário

Considerações similares às da seção 4.5.1 foram feitas para o texto do questionário, para o qual apresentamos a seguinte tabela com marcadores culturais em ordem alfabética:

Tabela 16 – Marcadores culturais presentes no questionário original, com frequências, e respectivas traduções, com frequências e porcentagens de ocorrência no questionário traduzido para o inglês

Marcador cultural (TO)	Frequência (TO)	Tradução (TT)	Frequência (TT)	Percentual (TT)
açudes	1	omissão do marcador	0	100,0%
carteira (de trabalho assinada)	2	formal contract	2	100,0%
cartório	1	registry office	1	100,0%
estado	5*	state	4	80,0%
		omissão do marcador	1	20,0%
igarapés	1	creeks	1	100,0%
indígena	4	indigenous	4	100,0%
município	19	municipality	19	100,0%
supletivo	4	omissão do marcador	4	100,0%

*Valor diferente do observado na lista de palavras porque uma das ocorrências refere-se a ‘estado civil’, que está fora do nosso foco de interesse.

Uma breve observação da tabela acima deixa clara a significativa diferença em relação aos marcadores retirados dos *releases* quanto à variedade de opções tradutórias. Por ser um texto mais fechado, cujas perguntas estão inter-relacionadas, acreditamos ter sido deliberada essa escolha. Se, no texto corrido dos *releases* há

maior liberdade para flexibilizar as traduções, este procedimento seria inapropriado no questionário.

A seguir, alguns comentários e exemplos acerca de cada escolha.

Açudes

No questionário em inglês a única referência a ‘açudes’ foi omitida, e assim o quesito traduzido assumiu um caráter mais genérico do que no original em português:

TO: 7 - Rios, açudes, lagos e igarapés

TT: 7 - *Rivers, lakes and creeks*

Carteira

Aqui destaca-se a expressão ‘carteira de trabalho assinada’, e não somente o substantivo ‘carteira’, ou seja, o significado concentra-se na relação de trabalho e não no documento propriamente dito. A relevância deste aspecto é mantida, assim, com a modulação para ‘*formal contract*’. A referência ao caráter empregatício do contrato fica evidenciada no início da expressão em inglês, em que a opção tradutória para o item ‘empregado com carteira de trabalho assinada’ apresenta-se como ‘*employee with a formal contract*’.

TO: 6.48 - Nesse trabalho era: 01 - Empregado com carteira de trabalho assinada

TT: 6.48 - *Your status in this job was: 01 - Employee with a formal contract*

Cartório

Apesar de haver, em países falantes do inglês, serviços similares aos prestados por nossos cartórios, a emissão e validação de registros e outros documentos é feita de maneira diversa; uma vez que a atuação dos cartórios está diretamente ligada ao contexto jurídico de nosso país, esta palavra pode ser considerada um marcador cultural.

Neste caso optou-se por uma tradução bastante genérica, por modulação: ‘*registry office*’. Uma consulta a glossários sobre a temática legal em países falantes da língua inglesa nos mostra que tal expressão remete ao uso em inglês britânico, segundo o dicionário *Merriam Webster Online*, que traz a definição: “*a place where marriages are performed and where records of births, marriages, and deaths*

are kept". De fato, no questionário do Censo 'cartório' é usado como referência ao local de registro de nascimento do cidadão:

TO: 6.13 - Tem registro de nascimento: 01 - Do cartório.

TT: 6.13 - *Has your birth been registered? 01 - Yes, as a birth certificate at registry office.*

Vale observar que, neste caso específico, a estrutura da pergunta foi bastante modificada. Se em português o foco é o respondente, na tradução o elemento central é o próprio nascimento e seu registro, e tal modificação resulta em um texto mais rebuscado, que contraria a tendência observada na maioria dos exemplos retirado dos questionário traduzido.

Para 'cartório', o professor Donal Hunter considerou a escolha genérica, mas também neutra, um aspecto positivo dadas as variações possíveis nos diversos países falantes da língua inglesa.

Estado

Classificação político-administrativa de um espaço que engloba diversos municípios e refere-se à subdivisão geográfica e administrativa do território brasileiro em 27 unidades que possuem governo próprio, embora não totalmente independente da administração nacional. No questionário do Censo, para facilitar a condução da pesquisa, aparece como uma opção (em parênteses) nos quesitos que incluem "Unidade da Federação", nome técnico adotado pelo IBGE, mas não de conhecimento do grande público.

A tradução literal '*state*' foi amplamente utilizada como equivalente de 'estado'. Vale a pena ressaltar o vocábulo em português aparece no questionário sempre em parênteses, após 'Unidade da Federação (UF)'. Tal recurso é adotado nas perguntas para facilitar a condução da pesquisa, uma vez que 'estado' parece ser de compreensão mais geral do que a opção específica UF. Essa relação é mantida no texto traduzido:

TO: 6.22 - Qual é a unidade da federação (estado) ou país estrangeiro de nascimento?

TT: 6.22 - *What is your federation unit (state) or foreign country of birth?*

Segundo o professor consultado, ‘states’ é uma opção adequada neste caso.

Igarapés

A tradução escolhida para esta única ocorrência foi ‘*creeks*’:

TO: 7 - Rios, açudes, lagos e igarapés

TT: 7 - *Rivers, lakes and creeks*

É importante ressaltar que essa escolha mantém uma unidade com o vocabulário adotado nos textos dos *releases*.

Indígena

No questionário, todas as referências a ‘indígenas’ em português referem-se à forma adjetiva, e receberam, na totalidade das traduções, a opção ‘*indigenous*’, de forma contrária aos *releases*, em que houve predominância de ‘*Indian*’:

TO: 6.10 - Qual (is)? (especifique a(s) língua(s) indígena(s) - até dois registros)

TT: 6.10 - *Specify the indigenous language(s) spoken (up to two entries)*

O quesito em questão ocorre como sequência ao item que indaga sobre o uso de língua indígena no domicílio. A tradução se apresenta de forma mais objetiva do que o original, pois dispensa o pronome interrogativo “Qual” e trata diretamente dos tipos de língua faladas.

Município

A tradução literal ‘*municipality*’ foi adotada ao longo de todo o questionário, tanto nos quesitos e pergunta quanto nas opções de resposta.

TO: 6.18 - Nasceu neste município?

TT: 6.18 - *Were you born in this municipality?*

Para este marcador também houve unidade com relação à tradução adotada nos *releases*.

Supletivo

Refere-se a uma modalidade de ensino de conteúdo condensado oferecida para adultos que querem completar o ciclo básico da educação; por ser específico de nosso sistema de educação, foi listado como marcador cultural.

Nas questões relativas à educação e série escolar a tradução mescla os conceitos do chamado EJA (educação de jovens e adultos), em inglês, '*youth and adult education*', e do ensino fundamental e médio. Perdeu-se, assim, o conteúdo relativo ao significado específico do 'supletivo' como modalidade condensada de educação.

TO: 06 - Educação de jovens e adultos - EJA – ou supletivo do ensino fundamental

TT: 06 - *Youth and adult basic education*

TO: 08 - Educação de jovens e adultos - EJA – ou supletivo do ensino médio

TT: 08 - *Youth and adult upper secondary education*

Neste caso o consultor também mencionou a dificuldade de se chegar a uma única tradução, uma vez que há diferentes modelos educacionais nos países com inglês nativo. Segundo o professor, não haveria um equivalente tradutório que ao mesmo tempo reúna os aspectos educativo, educacional e de otimização de tempo, como no caso de 'supletivo'. Uma sugestão genérica foi '*fast-track education*'.

Uma lista com todas as opções tradutórias alternativas sugeridas pelo consultor está disponível no Apêndice C.

4.6 Análise das características da tradução

Aqui apresentaremos comentários e exemplo acerca de ocorrências das características da tradução elencadas por Baker (1996), a saber: simplificação, explicitação e normalização. Tais características são aqui observadas apenas em termos da tradução dos marcadores culturais propriamente dita, sendo

desconsideradas, para fins deste trabalho, as demais numerosas ocorrências encontradas em todo o texto traduzido.

É importante ressaltar, ainda, que os exemplos a seguir foram retirados apenas dos *releases*. A comparação entre os questionários em português e inglês mostra que, para este tipo de texto e dado nosso objeto específico, marcadores culturais, apenas a normalização foi observada. Não identificamos a presença de explicitações ou simplificações.

4.6.1 Ocorrências de explicitação

Uma das características indicativas da explicitação é o maior tamanho dos textos traduzidos em relação aos respectivos originais, o que ocorre em apenas um dos nossos subcorpora de estudo, o dos *releases*. Mesmo assim, conforme observado a partir da Tabela 1, a diferença entre o texto original e tradução, em termos de número de *tokens*, foi mínima: apenas 0.48%.

Os textos em inglês do Censo 2010 apresentam significativa presença de traduções literais (conforme as Tabelas 15 e 16). Os resultados estatísticos e os exemplos extraídos do alinhamento entre os textos originais e traduzidos evidenciam que o número de ocorrências de normalização e de simplificação superam com ampla margem os casos de explicitação para este corpus de estudo. Apresentamos alguns exemplos a seguir.

Explicitação - Exemplo 1:

TO: Em 2010, o país possuía 6.329 aglomerados subnormais (assentamentos irregulares conhecidos como favelas, invasões, grotas, baixadas, comunidades, vilas, ressacas, mocambos, palafitas, entre outros) em 323 dos 5.565 municípios brasileiros.

TT: *In 2010 Brazil had 6,329 subnormal agglomerates in 2010 – irregular settlements known as favelas (slums), invasões (invaded properties), grotas (slums in deep valleys), baixadas (slums in low-lands), comunidades (poor communities), vilas (slums in villages), ressacas (slums in backwaters), mocambos (type of shack) and palafitas (stilt houses), among others – in 323 of the 5,565 municipalities.*

No excerto acima os tradutores optaram por manter os vocábulos originais em português, e oferecer, em parênteses, uma possível tradução. Observa-se que foi usada a palavra ‘*slum*’ (mais comum em inglês), em uma tentativa de explicitar/explicar o tipo de moradia a que o trecho se refere: ‘*slums in deep valleys*’, ‘*slums in villages*’ e ‘*slums in backwaters*’.

Explicitação - Exemplo 2:

TO: No total da população ocupada do País, os empregados com carteira de trabalho assinada representavam 44,5%,

TT: *Considering the total employed population in the country, workers with a formal employment contract represented 44.5%.*

Neste caso, optou-se por modificar a natureza do substantivo. Enquanto ‘carteira (assinada)’ em português trata do documento propriamente dito, a tradução em inglês prioriza a natureza do contrato de trabalho celebrado nesses casos, de modo a informar o leitor estrangeiro sobre o *status* de formalização empregatícia que o documento confere ao trabalhador brasileiro.

4.6.2 Ocorrências de normalização

Neste subitem listamos algumas ocorrências de normalização identificadas ao longo do texto. As normalizações, em geral, refletem a tentativa dos tradutores de reproduzirem estruturas mais comumente ...

Normalização - Exemplo 1:

TO: Pela primeira vez, o Censo investigará a língua falada pelos indígenas residentes no território nacional. Além de fornecer um diagnóstico da diversidade linguística existente no país, o mapeamento das línguas indígenas possibilitará a implementação de políticas públicas voltadas ao reconhecimento, preservação e promoção dessas línguas.

TT: *For the first time, the Census will investigate the language spoken by Indians living in the national territory. At the same time it will show a panorama of the linguistic*

diversity in the country, the mapping of such languages will make it possible to establish public policies aimed at their recognition, preservation and promotion.

Neste exemplo temos o uso de ‘*such*’ para retomar o tema ‘língua indígena’ e evitar a provável repetição de ‘*Indians*’, já presente na oração anterior.

Normalização - Exemplo 2:

TO: A pirâmide etária dos indígenas residentes fora das terras indígenas indica baixa fecundidade e mortalidade. Já para os indígenas residentes nas terras, a pirâmide etária ainda é resultante de uma alta natalidade e mortalidade.

TT: *The age pyramid of Indians who lived out of their land points to low levels of fertility and mortality. On the other hand, considering Indians living in those areas, the age pyramid results from high birth and mortality rates.*

Aqui optou-se pelo uso do adjetivo possessivo ‘*their*’ para novamente evitar a repetição de ‘*Indians*’, uma vez que o vocábulo aparece outras duas vezes no mesmo período.

Normalização - Exemplo 3:

TO: O IBGE incluiu esse tema no questionário do Censo atendendo a demandas de estudiosos da questão indígena e após discussões com um grupo de especialistas e representantes da Fundação Nacional do Índio – Funai.

TT: *IBGE has included this topic in the Census questionnaire as a reply to demands from scholars who study Indian-related issues and after a discussion with a group of specialists and representatives of the Brazilian National Indian Foundation (Funai).*

O terceiro exemplo traz um tipo de expressão comum na língua inglesa, em que o adjetivo é formado por um substantivo e o particípio ‘*related*’, significando ‘relativo a’. Assim, para exprimir a ideia geral da ‘questão indígena’, a equipe escolheu ‘*Indian-related issues*’.

Normalização - Exemplo 4:

TO: Tempo de deslocamento para o trabalho era menor para moradores de aglomerados no Rio de Janeiro

TT: *Agglomerate residents in Rio de Janeiro spend less time commuting to work*

Aqui também temos uma modificação de substantivo para adjetivo, transposição que resulta em uma estrutura mais comum na língua inglesa, mais sintética e sem a presença da preposição, e que também ressalta a relevância do item ‘aglomerados’.

Normalização - Exemplo 5:

TO: Em Belém, uma das características dominantes é a grande extensão das áreas de aglomerado subnormal.

TT: *A relevant feature in Belém is the large extension of the areas of subnormal agglomerates.*

No exemplo 5, a modificação ocorreu com passage da forma singular para o plural, de modo a expressar em língua inglesa o aspecto de generalidade que descreve as ‘áreas de aglomerado’.

Normalização - Exemplo 6:

TO: A partir de hoje, as prefeituras terão 20 dias para apresentar suas avaliações sobre os números divulgados e, no dia 29 de novembro, o IBGE divulgará os resultados das populações dos municípios.

TT: *From today on, city halls will have 20 days to present their evaluations on the released figures, and, on November 29, IBGE will release results of municipal populations.*

Este último exemplo selecionado apresenta mais uma transposição de substantivo para adjetivo, também evitando a repetição da preposição e conferindo um caráter mais fluido ao texto.

4.6.3 Ocorrências de simplificação

Confirmando o que sugerem as estatísticas gerais (Tabela 1), que apresenta uma razão forma/item significativamente mais baixa para os *releases* em inglês, podemos encontrar no texto traduzido ocorrências de simplificação. Diversos

exemplos referem-se às mudanças na estrutura do texto, como alterações na pontuação e quebra de períodos.

No caso específico dos marcadores culturais, a simplificação se deu pela utilização de traduções mais genéricas para tais palavras, por conta da inexistência ou desconhecimento de opções tradutórias mais próximas ao contexto da língua inglesa. O trecho abaixo traz, juntamente com a tradução do vocábulo 'indígena' dois exemplos de simplificação.

Simplificação - Exemplo 1:

TO: No total do país, verificou-se a existência de esgoto a céu aberto e depósitos de lixo nos logradouros do entorno em 11,0% e 5,0% dos domicílios, respectivamente.

TT: *In the country as a whole, the survey registered the existence of open sewage and garbage accumulation in the surroundings at 11.0% and 5.0%, respectively.*

O exemplo acima tem como contexto a descrição do entorno dos domicílios. O texto original em português traz 'logradouros do entorno', enquanto o texto inglês apresenta somente 'surroundings' e omite a tradução para logradouros, assim simplificando o conteúdo do texto como um todo.

Este caso específico exemplifica os resultados de razão forma item mais baixa, uma vez que no texto em português temos duas opções diferenciadas, 'endereço/endereços' e 'logradouro/logradouros', enquanto nos textos em inglês, conforme observado na Tabela 15, predomina a tradução 'address/addresses', ou a completa omissão da palavra.

Simplificação - Exemplo 2:

TO: Para conhecer o perfil das famílias que não registram os filhos e onde elas moram, o IBGE perguntará, pela primeira vez, se as crianças com até 10 anos possuem registro de nascimento ou Registro Administrativo de Nascimento Indígena, no caso das crianças indígenas.

TT: *In order to know the profile of these families and where they live, IBGE will ask, for the first time, if children aged up to 10 have a birth certificate or a correspondent document for indigenous children, if it is the case.*

O exemplo 2 traz uma generalização na referência ao documento RANI, Registro Administrativo de Nascimento Indígena, em que não somente o marcador ‘indígena’ é omitido, mas também todo o nome do documento, sendo mencionado apenas como ‘*correspondent document*’ em alusão à certidão de nascimento dos não-indígenas.

Observe-se também o equivocado uso de ‘correspondent’, para ‘correspondente’, quando a opção desejada provavelmente seria “*corresponding*”.

Simplificação - Exemplo 3:

TO: Na região Centro-Oeste, havia a menor incidência de domicílios situados em logradouros com depósito de lixo (3,7%) e esgoto a céu aberto (2,9%).

TT: *In the Central-West region, there was the lowest incidence of households settled in areas with garbage accumulation (3.7%) and open sewage (2.9%).*

Mais uma vez a tradução de ‘logradouros’, que aqui aparece em uma generalização, tendo sido traduzidos como ‘*areas*’, opção tradutória já adotada para outros vocábulos do texto em português (áreas e regiões). Este exemplo também corrobora os dados de menor riqueza vocabular do texto traduzido, como indicativo de simplificação.

4.7 Comparação com textos originalmente escritos em inglês

Nesta seção fazemos um comparativo entre textos sobre o tema Censo Demográfico retirados de fontes de dois países falantes do inglês: os Estados Unidos da América e o Canadá. Por razões de similaridade temática, e também de proximidade de data do evento ao qual o material se refere, os corpora de comparação foram obtidos a partir de textos do *US Census Bureau* para a edição de 2010 do censo americano; e do *Statistics Canada* (para a edição de 2011 do censo do Canadá).

De modo a melhor comparar o conteúdo do material, analisamos as ocorrências mais relevantes presentes nas palavras-chave, exibidas abaixo, para os textos em inglês – *releases* e descritivos de resultados do censo – obtidas por meio do *WordSmith Tools*, para os três institutos:

Tabela 17 – Trinta primeiras palavras-chave dos textos em inglês, publicados pelo *US Census Bureau*, *Statistics Canada* e IBGE, em ordem decrescente de chavacidade

	US Census Bureau	Log_L	Statcan	Log_L	IBGE	Log_L
1	<i>census</i>	7.084,57	<i>population</i>	5.487,28	<i>census</i>	3.808,74
2	<i>population</i>	2.928,33	<i>census</i>	5.306,98	<i>IBGE</i>	2.919,40
3	<i>bureau</i>	1.683,34	<i>Canada</i>	4.19263	<i>population</i>	2.817,93
4	<i>households</i>	704,64	<i>income</i>	2.734,30	<i>persons</i>	2.713,75
5	<i>race</i>	653,09	<i>aboriginal</i>	2.322,47	<i>municipalities</i>	2.265,31
6	<i>mail</i>	634,12	<i>households</i>	2.113,31	<i>households</i>	1.795,76
7	<i>Hispanic</i>	511,12	<i>household</i>	1.302,61	<i>agglomerates</i>	1.665,34
8	<i>states</i>	507,40	<i>proportion</i>	1.286,63	<i>percentage</i>	1.450,89
9	<i>Asian</i>	468,07	<i>provinces</i>	1.176,66	<i>northeast</i>	1.991,34
10	<i>areas</i>	359,11	<i>statistics</i>	1.176,54	<i>income</i>	1.117,42
11	<i>groves</i>	358,59	<i>age</i>	1.169,57	<i>units</i>	1.108,71
12	<i>participation</i>	354,67	<i>percentage</i>	1.157,39	<i>Rio</i>	998,22
13	<i>black</i>	319,01	<i>immigrants</i>	1.134,96	<i>southeast</i>	949,10
14	<i>native</i>	317,20	<i>Inuit</i>	1.132,81	<i>household</i>	938,13
15	<i>nation</i>	305,54	<i>Quebec</i>	1.061,61	<i>Brazil</i>	936,92
16	<i>data</i>	304,69	<i>Alberta</i>	1.057,00	<i>areas</i>	883,72
17	<i>percentage</i>	291,63	<i>families</i>	1.023,42	<i>region</i>	824,13
18	<i>people</i>	282,48	<i>children</i>	1,022.44	<i>Brazilian</i>	819,39
19	<i>respondents</i>	272,42	<i>territories</i>	983,54	<i>Janeiro</i>	793,59
20	<i>Alaska</i>	265,47	<i>couples</i>	955,62	<i>enumerators</i>	773,62
21	<i>populations</i>	265,26	<i>Ontario</i>	931,14	<i>age</i>	740,27
22	<i>forms</i>	259,84	<i>metropolitan</i>	886,26	<i>Indian</i>	683,04
23	<i>racess</i>	247,14	<i>Métis</i>	884,80	<i>Indians</i>	673,30
24	<i>counties</i>	243,86	<i>seniors</i>	876,97	<i>housing</i>	672,53
25	<i>rates</i>	228,63	<i>language</i>	757,40	<i>municipality</i>	642,52
26	<i>campaign</i>	225,31	<i>Canadian</i>	753,50	<i>residents</i>	622,41
27	<i>growth</i>	225,06	<i>catalogue</i>	744,59	<i>percentages</i>	584,06
28	<i>rate</i>	217,12	<i>French</i>	743,74	<i>schooling</i>	557,61
29	<i>partners</i>	204,69	<i>Columbia</i>	711,64	<i>color</i>	534,41
30	<i>increase</i>	202,23	<i>English</i>	689,73	<i>fertility</i>	515,79

A tabela acima evidencia a presença de três ocorrências idênticas nas três listas, simultaneamente: ‘*census*’, ‘*population*’ e ‘*household(s)*’, que se referem ao tipo de investigação realizada, ao seu público alvo (e também fornecedor de informações), e à unidade de coleta dos dados. É possível observar que tais ocorrência encontram-se em uma posição próxima ao topo da lista, ou seja, que apresentam alto índice de chavicidade nos três casos.

Observa-se também a ocorrência de palavras do mesmo campo semântico, por exemplo, aquelas que se referem às populações nativas, como ‘*native*’, ‘*Inuit*’, ‘*aboriginal*’ e ‘*Indian(s)*’, conforme mostrado abaixo:

US Census Bureau:

Overall, 1.2 million people, or 0.4 percent of all people in the United States, identified as native Hawaiian and other Pacific islander (nhpi), either alone or in combination with one or more races.

Statistics Canada:

In 2011, 59,445 people identified as Inuit. They represented 4.2% of the total Aboriginal population and 0.2% of the total Canadian population.

IBGE:

For the first time, the Census will investigate the language spoken by Indians living in the national territory.

Outro exemplo é a presença de topônimos, como os nomes dos próprios países, tais como *Brazil* e *Canada*, e de localidades como *Rio*, *Alaska*, *Quebec* e *Alberta*. Vale a pena mencionar que, embora haja coincidência nas palavras-chave, nenhum dos marcadores culturais traduzidos nos textos do Censo 2010 em inglês se encontra presente nos textos originalmente escritos em inglês. Isso parece confirmar a especificidade do tema quanto à cultural e realidade local de cada país, uma vez que, conforme já tratado em seções anteriores, os marcadores dos textos do corpus representam elementos bastante peculiares do Brasil.

Com relação aos questionários, apresentaremos no quadro a seguir apenas as quinze primeiras palavras-chave, uma vez que foram geradas para o questionário do *Statistics Canada* somente quinze ocorrências.

Tabela 18 – Quinze primeiras palavras-chave dos questionários em inglês, publicados pelo *US Census Bureau*, *Statistics Canada* e IBGE, em ordem decrescente de chavacidade

	US Census Bureau	Log_L	Statcan	Log_L	IBGE	Log_L
1	<i>person</i>	1.102,87	<i>persons</i>	149,61	<i>municipality</i>	314,30
2	<i>Hispanic</i>	369,66	<i>census</i>	137,80	<i>unit</i>	245,90
3	<i>name</i>	284,48	<i>questionnaire</i>	134,70	<i>housing</i>	179,18
4	<i>question</i>	247,62	<i>residence</i>	113,42	<i>July</i>	176,22
5	<i>latino</i>	231,00	<i>Canada</i>	103,52	<i>federation</i>	148,69
6	<i>origin</i>	206,77	<i>address</i>	92,62	<i>resident</i>	93,94
7	<i>apartment</i>	205,13	<i>person</i>	86,25	<i>residents</i>	92,52
8	<i>mobile</i>	191,61	<i>step</i>	78,13	<i>country</i>	92,01
9	<i>race</i>	187,99	<i>information</i>	58,64	<i>education</i>	75,43
10	<i>months</i>	171,83	<i>page</i>	48,35	<i>person</i>	65,89
11	<i>box</i>	135,42	<i>statistics</i>	47,20	<i>indigenous</i>	64,24
12	<i>puerto</i>	128,99	<i>question</i>	44,11	<i>birth</i>	57,13
13	<i>Condominium</i>	108,89	<i>name</i>	35,07	<i>difficulty</i>	53,70
14	<i>Spanish</i>	105,78	<i>online</i>	29,47	<i>nursery</i>	46,02
15	<i>home</i>	105,01	<i>telephone</i>	28,69	<i>household</i>	44,98

Os resultados acima apresentam um maior nível de heterogeneidade do que a comparação entre os *releases* e os textos referentes aos censos americano e canadense. A única ocorrência de fato comum entre as três listas é o vocábulo ‘*person/persons*’, que se refere aos próprios moradores do domicílio, ou unidade de coleta, nos três contextos, conforme exemplos abaixo:

US Census Bureau:

Person 1 is the person living or staying here in whose name this house or apartment is owned, being bought, or rented. If there is no such person, start with the name of any adult living or staying here.

Statistics Canada:

If more than six persons live here, you will need an extra questionnaire; call 1-877-777-2011.

IBGE:

3.01 - *Was any person who used to live with you living in another country on July 31, 2010?*

Observamos diversas ocorrências pertencentes a um mesmo campo semântico, por exemplo, relacionadas à moradia: *'apartment'*, *'condomminium'* e *'home'* (US Census Bureau); *'residence'* e *'address'* (Statistics Canada); *unit*, *housing* e *'household'* (IBGE). Um outro exemplo refere-se as questões raciais/étnicas: *'hispanic'*, *'latino'* e *'Spanish'* (US Census Bureau) e *'indigenous'* (IBGE).

As diferenças mais presentes neste caso podem indicar que, conforme mencionado anteriormente, o questionário é um texto de grande especificidade. Além de abordar o tema, ele reflete um tipo de tratamento metodológico para os dados em cada um dos censos, com peculiaridades observáveis já na coleta das informações. Desse modo, a lista de palavras é bastante condizente com essa particularidade, uma vez que os três questionários abordam de forma diferente o público-alvo da investigação.

CONSIDERAÇÕES FINAIS

Nesta dissertação, buscamos investigar a tradução de marcadores culturais selecionados de textos do Censo 2010 – dos tipos *release* e questionário – em traduções de português para o inglês disponibilizadas no *website* do Instituto Brasileiro de Geografia e Estatística (IBGE). Investigamos também a ocorrência de características da linguagem da tradução, a saber: explicitação, normalização e simplificação, e fizemos comparações entre o vocabulário mais relevante nos textos provenientes do IBGE e de outros dois órgãos de mesma natureza, o *Statistics Canada* e o *US Census Bureau*.

Visando ao enriquecimento da discussão, também recorremos a um professor nativo da língua inglesa – não especialista em estatística ou área correlata ao trabalho do IBGE – para ler os textos e discutir aspectos das traduções, assim como para apresentar opções outras. Uma vez que nosso trabalho não possui um caráter prescritivo, mas sim descritivo, a ideia foi ter o olhar nativo de um “leitor comum”, e compilar suas observações quando em contato com os textos estudados.

A relevância desta pesquisa evidencia-se em dois aspectos principais: a importância e a popularidade da temática dos textos, uma vez que o Censo Demográfico revela aspectos importantes do país e de seu povo; e a quase total ausência de estudos acadêmicos acerca das traduções feitas por órgãos governamentais em nosso país. Assim, ao apresentar um estudo que contempla o IBGE e a tradução, tivemos a oportunidade de dar um pouco de visibilidade à produção do Instituto e de seus profissionais tradutores.

Em termos dos resultados propriamente ditos, o tratamento dos *releases* por meio do programa *WordSmith Tools* revelou, já no início da pesquisa, alguns aspectos peculiares desse material. Os marcadores culturais selecionados segundo os domínios da cultura ecológica, ideológica, material e social foram traduzidos por meio das modalidades de tradução literal, transposição, modulação, empréstimo, ocorrendo também a omissão, em diversos casos. Em outros, houve inclusive a sobreposição de modalidades para uma mesma opção tradutória.

Os textos do questionário, ao contrário, tiveram menor variação entre as modalidades, sendo predominantes as do tipo tradução literal e modulação. Em geral, as maiores diferenças quanto à tradução dos textos analisados resultam de suas

próprias características intrínsecas. Os *releases* são mais longos, elaborados por jornalistas e destinados ao provimento de informações para o grande público. O questionário, por sua vez, é elaborado por especialistas técnicos, é extremamente específico – de modo a impossibilitar interpretações dúbias –, devendo soar claro quando lido para o entrevistado.

De forma geral, observa-se no questionário uma tradução mais simplificadora do que no caso dos *releases*. Dada a necessidade de um texto bastante enxuto e sem espaço para questionamentos por parte do entrevistado, a tradução em inglês do questionário, na maioria dos casos apresentados, trouxe opções mais literais e genéricas. Ressalte-se também que tal tendência pode estar relacionada ao objetivo de cada tipo de tradução. Embora ambos os textos estejam disponibilizados para o público de língua estrangeira, os *releases* teriam o objetivo de informar, situar o leitor autóctone acerca dos resultados mais relevantes de toda a operação censitária. O questionário, ao contrário, é parte de uma etapa metodológica que interessaria mais aos especialistas da área, não ao grande público. Além disso, por basear-se na realidade do Brasil, tal questionário provavelmente não teria aplicação real em contextos internacionais, a não ser para referência e consulta.

O domínio cultural mais observado em ambos os casos foi o da cultura social. Tal resultado está em conformidade com a própria natureza do Censo, que é classificado dentro do IBGE como uma pesquisa de caráter social/populacional. Os demais tipos de marcadores também foram encontrados no corpus, mas destacamos, com apenas duas ocorrências nos *releases* e nenhuma no questionário, os marcadores pertencentes ao domínio da cultura ideológica. Os dois marcadores presentes nos *releases* provêm das repostas dadas pelo entrevistado ao recenseador, uma vez que o quesito sobre religião está formulado no questionário como uma pergunta aberta, ou seja, que não apresenta opções pré-definidas.

Conforme sugerido pelas estatísticas gerais, que mostram bastante proximidade de tamanho (número de *tokens*) entre os TO e TT, observou-se que a explicitação foi característica da linguagem de tradução menos observada entre os marcadores culturais nos *releases*, enquanto a normalização foi a que mais ocorreu. Para o questionário não foram encontrados exemplos numerosos do uso da linguagem de tradução, provavelmente por conta das razões intrínsecas já mencionadas.

A comparação com textos dos organismos estatísticos estrangeiros revelou uma maior similaridade entre os três textos discursivos, com coincidência de três palavras entre as 30 de maior chavicidade. Ainda neste caso, observamos também significativas ocorrências de topônimos, que foram desconsiderados para fins desta pesquisa.

Conforme o esperado, dada a especificidade dos questionários, este tipo de texto apresentou menor similaridade na comparação entre os três institutos, sendo apenas comum a ocorrência de palavras do mesmo campo semântico. Um ponto que merece destaque é a ausência dos marcadores presentes nos TT do IBGE em quaisquer dos textos originalmente escritos em inglês, o que acaba por reforçar o caráter de cultura local presente em tais vocábulos.

Dadas as possibilidades diversas de investigação segundo a metodologia aqui apresentada, com base na Linguística de Corpus, acreditamos ser possível realizar um estudo ainda mais abrangente acerca das traduções dos textos produzidos pelo IBGE – com temáticas afins às do Censo 2010 ou outras pertinentes ao Instituto. Uma breve consulta ao *website* da fundação apresenta ao pesquisador uma gama de possíveis temas para estudo, todos relacionados ao desafio constante da equipe “ibgeana” de tradução: ser capaz de comunicar conteúdos extremamente brasileiros aos usuários que buscarão informações em inglês – um público heterogêneo e que está espalhado ao redor do mundo.

Assim, ao fazermos aqui considerações sobre a prática tradutória nesse contexto, talvez consigamos dar à sociedade uma contribuição adicional a partir das informações produzidas pelo IBGE, não mais tendo como base apenas dados numéricos e frios, mas valorizando a palavra para gerar reflexão e estimular a produção de novos conhecimentos a partir do brasileiríssimo trabalho do Instituto.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Textos do Corpus

IBGE, **2010 Population Census**. Disponível em:
<<http://www.ibge.gov.br/english/estatistica/populacao/censo2010/default.shtm>>.
Acesso em: 9 jul. 2016.

IBGE, **Censo Demográfico 2010**. Disponível em:
<<http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm>>.
Acesso em: 9 jul. 2016.

STATCAN, **Aboriginal Peoples in Canada: First Nations People, Métis and Inuit 2011**. Disponível em: <<https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-011-x/99-011-x2011001-eng.cfm>>. Acesso em: 05 mai. 2018.

STATCAN, **Education and occupation of high-income Canadians**. Disponível em:
<https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-014-x/99-014-x2011003_2-eng.cfm>. Acesso em: 05 mai. 2018.

STATCAN, **Education in Canada: Attainment, Field of Study and Location of Study**. Disponível em: <<https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/99-012-x2011001-eng.cfm>>. Acesso em: 05 mai. 2018.

STATCAN, **Homeownership and Shelter Costs in Canada**. Disponível em:
<<http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-014-x/99-014-x2011002-eng.cfm>>. Acesso em: 05 mai. 2018.

STATCAN, **Immigration and Ethnocultural Diversity in Canada**. Disponível em:
<<http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011001-eng.pdf>>. Acesso em: 05 mai. 2018.

STATCAN, **Linguistic Characteristics of Canadians**. Disponível em:
<<http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-314-x/98-314-x2011001-eng.cfm>>. Acesso em: 05 mai. 2018.

STATCAN, **Living arrangements of seniors**. Disponível em:
<https://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-312-x/98-312-x2011003_4-eng.cfm>. Acesso em: 05 mai. 2018.

STATCAN, **Portrait of Canada's Labour Force**. Disponível em:
<<https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/99-012-x2011002-eng.cfm>> Acesso em: 05 mai. 2018.

STATCAN, **Portrait of Families and Living Arrangements in Canada**. Disponível em: <<https://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-312-x/98-312-x2011001-eng.cfm>>. Acesso em: 05 mai. 2018.

STATCAN, **The Canadian Population in 2011: Age and Sex**. Disponível em:

<<http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-311-x/98-311-x2011001-eng.cfm?wbdisable=true>>. Acesso em: 05 mai. 2018.

STATCAN, **The Canadian Population in 2011: Population Counts and Growth**. Disponível em: <<https://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-310-x/98-310-x2011001-eng.cfm>>. Acesso em: 05 mai. 2018.

UNSD, **Long-Form Questionnaire (2010 Brazilian Census)**. Disponível em: <<https://unstats.un.org/unsd/demographic/sources/census/quest/BRA2010enl.pdf>>. Acesso em: 15 jan. 2017.

UNSD, **Questionário da Amostra (Censo 2010)**. Disponível em: <<https://unstats.un.org/unsd/demographic/sources/census/quest/BRA2010ptl.pdf>>. Acesso em: 15 jan. 2017.

US CENSUS BUREAU, **News Archive 2012**. Disponível em: <<https://www.census.gov/newsroom/releases/archives/2012.html>>. Acesso em: 10 jun. 2018.

US CENSUS BUREAU, **News Archive 2011**. Disponível em: <<https://www.census.gov/newsroom/releases/archives/2011.html>>. Acesso em: 10 jun. 2018.

US CENSUS BUREAU, **News Archive 2010**. Disponível em: <<https://www.census.gov/newsroom/releases/archives/2010.html>>. Acesso em: 10 jun. 2018.

2. Dicionários e glossários

Dicionário Michaelis Online. Disponível em: <<https://michaelis.uol.com.br/>>. Acesso em: 03 ago. 2018.

Glossário Online do Censo 2010. Disponível em: <https://censo2010.ibge.gov.br/apps/atlas/pdf/209_213_Glossario_ATLASDEMO%202010.pdf>. Acesso em: 15 jul. 2018.

Merriam-Webster Online. Disponível em: <<https://www.merriam-webster.com/>>. Acesso em: 15 jul. 2018.

Povos Indígenas no Brasil. Disponível em: <<https://pib.socioambiental.org/pt/Povo:Munduruku>>. Acesso em 15 jul. 2018.

3. Obras da base teórica e metodológica

AUBERT, Francis Henrik. **A tradução do intraduzível**, São Paulo: FFLCH/ USP, 1981.

_____. Modalidades de tradução: teoria e resultados. In: **TradTerm**, 5.1. São Paulo: CITRAT/FFLCH/USP, 1998. p. 99-128.

_____. Traduzindo as diferenças extra-linguísticas – procedimentos e condicionantes. In: **TradTerm**, 5.1. São Paulo: CITRAT/FFLCH/USP, 2003, p. 151-172.

_____. Indagações acerca dos marcadores culturais na tradução. **Revista de Estudos Orientais**, São Paulo: DLO/FFLCH/USP, v5, 2006, p. 23-36.

BAKER, M. Corpus Linguistics and Translation Studies: Implications and Applications. In: _____.; FRANCIS, G. TOGNINI-BONELLI, E. **Text and Technology: In Honour of John Sinclair**. Amsterdam; Philadelphia: John Benjamins Publishing, 1993, p. 233-250.

_____. Corpora in translation studies: an overview and some suggestions for future research. **Target**, v. 7, n. 2, 1995, p. 223-243.

_____. Corpus-based translation studies: the challenges that lie ahead. In: SOMERS, H. (Ed). **Terminology, LSP and translation studies in language engineering in honour of Juan C. Sager**. Amsterdam: John Benjamins, 1996. p. 177-186.

_____. SALDANHA, G. (eds), **Routledge Encyclopedia of Translation Studies**. 2nd Edition, London / New York: Routledge, 2009.

BARROS, L. A. **Curso Básico de Terminologia**. São Paulo: Edusp, 2004.

BERBER SARDINHA, T. BERBER SARDINHA, A. P. *Linguística de corpus: histórico e problemática*. **DELTA: documentação de estudos em linguística teórica e aplicada**, São Paulo, v. 16, n. 2, 2000, p. 323-367.

_____. **Linguística de Corpus**. São Paulo: Editora Manole Ltda, 2004.

_____. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

CAMARGO, D. C. **Padrões de estilo de tradutores: um estudo de semelhanças e diferenças em corpora de traduções literárias, especializadas e juramentadas**. 2005. 512 f. Tese (Livre-Docência em Estudos da Tradução) - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto. 2005.

_____. **Metodologia de pesquisa em Tradução e Linguística de Corpus**. São Paulo São José do Rio Preto: Cultura Acadêmica/Laboratório Editorial do IBILCE, UNESP, 2007.

CERQUEIRA, C. A., GIVISIEZ, G. H. N. Conceitos básicos em demografia e dinâmica demográfica brasileira. In: **Introdução à Demografia da Educação** (ABEP), 2004, p. 15-44. Disponível em: <<http://www.abep.nepo.unicamp.br/docs/outraspub/demoedu/parte1cap1p13a44.pdf>>. Acesso em mar. 2017.

CHESTERMAN, A. Proposal for a Hyeronymic Oath. In: PYM, Anthony. **The Return to ethics**. Special issue of the *Translator* 7, Manchester: St. Jerome Publishing, 2001, p 139-154.

EVEN-ZOHAR, I. The position of translated literature within the literary polysystem. In: VENUTI, L. (Ed.). **The translation studies reader**. London: Routledge, 2000, p. 192-198.

HOLMES, J. S. The Name and Nature of Translation Studies. In: VENUTI, L. **The Translation Studies Reader**. London: Routledge Press, 2000, p. 172-186.

LAVIOSA, S. **Corpus-based Translation Studies: Theory, Findings, Applications**. Amsterdam/Athanta: Rodopi, 2002.

MUNDAY, Jeremy. **Introducing translation studies: theories and applications**. Londres: Routledge, 2001.

OLIVEIRA, L. A. P., SIMÕES, C. C. S. O IBGE e as Pesquisas Populacionais. **Revista Brasileira de Estudos Populacionais**. v. 22, n. 2, jul-dez 2005, p. 291-302. Disponível em: <http://www.scielo.br/pdf/rbepop/v22n2/v22n2a06.pdf>. Acesso em mar. 2017.

PYM, A. The Return to ethics in Translation Studies. In: PYM, A. **The Return to ethics**. Special issue of the *Translator* 7, Manchester: St. Jerome Publishing, 2001, p 129-138.

RÓNAI, P. Problemas Gerais da Tradução. In: PORTINHO, W.M. (Org.). **A Tradução Técnica e seus Problemas**. São Paulo: Álamô, 1983, p. 1-16.

_____. **Escola de Tradutores**. São Paulo: José Olympio Editora, 2012.

SINCLAIR, J. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

SOARES, W. M. **Tradução Técnica e Marcadores Culturais: o Caso de uma Transposição Didática**. 2013. 63 f. Projeto Final (Graduação em Letras/Tradução). Universidade de Brasília (UNB), 2013.

TAGNIN, S. E. O. Linguística de Corpus na e para a Tradução. In: VIANA, Vander e _____, S. E. O. (Org.). **Corpora na Tradução**. 1. ed. São Paulo: HUB, 2015, p.19-56.

TOGNINI-BONELLI, E. **Corpus Linguistics at Work**. Amsterdam/Philadelphia: John Benjamins Press, 2001.

TOURY, G. **Descriptive translation studies and beyond**. Amsterdam: John Benjamins Press, 1995.

ZANETTE, R. I. C. O Paraná no guia turístico: uma análise de marcadores culturais. **Revista Línguas & Letras**, v. 17, 2016, p. 251-267.

ZAVAGLIA, A. As relações culturais na tradução de textos especializados. **Linguasagem**, v. 10, 2009, p. 1-9.

ZYNGIER, Sonia, Viana, Vander & Silveira, Natália Giordani. 2011. Discurso literário e linguística de *corpus*: uma visão empírica. **Cadernos de Letras** (UFRJ) n.28 - jul. 2011, p. 99-107.

**APÊNDICE A – Lista de topônimos encontrados nos *releases* em português,
em ordem decrescente de frequência e classificados por tipo**

Topônimo	Frequência	Tipo
Brasil	211	País
Norte	130	Grande Região
Nordeste	107	Grande Região
Sudeste	96	Grande Região
Rio de Janeiro	68	Unidade da Federação
São Paulo	61	Unidade da Federação
Sul	54	Grande Região
Santa Catarina	27	Unidade da Federação
Minas Gerais	25	Unidade da Federação
Distrito Federal	22	Território autônomo
Maranhão	22	Unidade da Federação
Paraná	20	Unidade da Federação
Rio Grande do Sul	20	Unidade da Federação
Acre	19	Unidade da Federação
Roraima	19	Unidade da Federação
Amapá	18	Unidade da Federação
Amazonas	18	Unidade da Federação
Bahia	15	Unidade da Federação
Belo Horizonte	15	Município
Ceará	12	Unidade da Federação
Rondônia	12	Unidade da Federação
Estados Unidos	11	País
Goiânia	11	Município
Goiás	11	Unidade da Federação
Japão	10	País
Mato Grosso	10	Unidade da Federação
Salvador	10	Município
Macapá	9	Município
Pará	9	Unidade da Federação
Piauí	9	Unidade da Federação
Rio Grande do Norte	9	Unidade da Federação
Fortaleza	8	Unidade da Federação
Manaus	8	Município
Porto Alegre	8	Município
Recife	8	Município
Alagoas	7	Unidade da Federação
Belém	7	Município
Brasília	7	Capital /DF
Curitiba	7	Município
Espírito Santo	7	Unidade da Federação
Paraíba	7	Unidade da Federação
Pernambuco	7	Unidade da Federação

Florianópolis	6	Município
Maceió	5	Município
Mato Grosso do Sul	5	Unidade da Federação
México	5	País
Tocantins	5	Unidade da Federação
Argentina	4	País
Austrália	4	País
Balbinos	4	Município
Finlândia	4	País
Paraguai	4	País
Santos	4	Município
Sergipe	4	Município
Boa Vista	3	Município
Cabo Verde	3	País
Centro Oeste	3	Grande Região
Costa Rica	3	País
Espanha	3	País
Europa	3	Continente
Palmas	3	Município
Portugal	3	País
República Dominicana	3	País
Rio Branco	3	Município
Rocinha	3	Aglomerado subnormal
Teresina	3	Município
Venezuela	3	País
Américas	2	Continente
Ásia	2	Continente
Batovi	2	Terra indígena
Bolívia	2	País
Canadá	2	País
Centenário	2	Município
China	2	País
Coréia do Sul	2	País
Costa do Marfim	2	País
Filipinas	2	País
Cubatão	2	Município
Indonésia	2	País
Inglaterra	2	País
Mundo Verde/Cachoeirinha	2	Terra indígena
Panamá	2	País
Papua Nova Guiné	2	País
Paquistão	2	País
Porto Velho	2	Município
Rússia	2	País
São Luís	2	Município
Suíça	2	País
Tailândia	2	País

Vitória	2	Município
Zâmbia	2	País
Zo'E	2	Terra indígena
André da Rocha	1	Município
Angola	1	País
Araguainha	1	Município
Areões	1	Terra indígena
Aripuanã	1	Terra indígena
Arvoredo	1	Município
Ava Canoeiro	1	Terra indígena
Badjonkore	1	Terra indígena
Baía do Guató	1	Terra indígena
Bertópolis	1	Município
Boca do Acre	1	Município
Borá	1	Município
Casa Amarela	1	Município
Cedro do Abaeté	1	Município
Chile	1	País
Colômbia	1	País
Doutor Ulysses	1	Município
EUA	1	País
Fernandes Tourinho	1	Município
Fernando de Noronha	1	Município
Frei Gaspar	1	Município
Haiti	1	País
Itália	1	País
Itatinga	1	Terra indígena
Ipueiras	1	Município
Jaboatão dos Guararapes	1	Município
Jamaica	1	País
Lajeado Grande	1	Município
Lavandeira	1	Município
Lavínia	1	Município
Maraã Urubaxi	1	Município
Macuco	1	Município
Marrocos	1	País
Mapari	1	Terra indígena
Mateiros	1	Município
Miguel Leão	1	Município
Miravânia	1	Município
Moçambique	1	País
Monte Santo	1	Município
Nova Castilho	1	Município
Nova Holanda	1	Aglomerado subnormal
Nova Ramada	1	Município
Nova Zelândia	1	País
Olinda	1	Município

Oliveira de Fátima	1	Município
Paraisópolis	1	Aglomerado subnormal
Paraná do Paricá	1	Terra indígena
Paris	1	Município
Parque União/Nova Holanda	1	Aglomerado subnormal
Pedra Branca do Amapari	1	Município
Peru	1	País
Pinheiro Preto	1	Município
Ponte Alta do Bom Jesus	1	Município
Pracinha	1	Município
Reino Unido	1	Grupo de países
Rio da Conceição	1	Município
Rio das Ostras	1	Município
Riozinho do Alto Envira	1	Terra indígena
Rio Omerê	1	Terra indígena
Sagarana	1	Terra indígena
Santa Helena de Minas	1	Município
Santo Antônio do Rio Abaixo	1	Município
São Brás	1	Município
São Caetano do Sul	1	Município
São Gabriel da Cachoeira	1	Município
São Geraldo da Piedade	1	Município
São Tomé e Príncipe	1	País
Sepoti	1	Terra indígena
Sobralia	1	Município
Sorocaba	1	Município
Suécia	1	País
Tupirama	1	Município
Uru	1	Município
Uruguai	1	País
Vidigal	1	Aglomerado subnormal

APÊNDICE B – Marcadores culturais presentes nas *KeyWords* e *WordLists* dos releases e questionário não selecionados para análise

bairro	Palavra que usualmente designa uma subdivisão urbana dentro de um município; é aqui enquadrada como marcador cultural por não encontrar equivalente exato na língua inglesa, uma vez que tal tipo de subdivisão social e geográfica se altera de país para país
<i>Munduruku</i>	Grupo indígena brasileiro que ocupa terras no Pará, Amazonas e Mato Grosso.
oca(s)	Nome que designa a moradia indígena brasileira, que difere em formato e material das habitações de indígenas de outras localidades, como as dos índios norte-americanos, por exemplo
palafita(s)	Tipo de habitação de madeira construída sobre a água, e comum em beiras de rios, em especial no Brasil, Ásia e África
Pataxó	Povo indígena brasileiro que habita regiões da Bahia
pinguela	Tipo de ponte improvisada com tocos de madeira ou troncos de árvore, comumente utilizada para atravessar áreas de rio no interior do Brasil
pretas	Aqui a marca cultural refere-se não exclusivamente à cultura local, mas ao uso que o IBGE faz do vocabulário, uma vez que o instituto decidiu adotar, em seus estudos de cor e raça, a denominação preto (a) para a população descendente de negros africanos
Sateré Mawé	Povo indígena que atualmente habita a periferia urbana de Manaus, no estado do Amazonas
Tupinambá	Povo indígena que atualmente ocupa áreas da região Sudeste, incluindo centros urbanos
Xavante	Povo indígenas que ocupa 12 regiões localizadas em Mato Grosso e Goiás

Yanomami	Indígenas que habitam áreas de Roraima e Amazonas, no Brasil, e também regiões da Venezuela
----------	---

APÊNDICE C – Marcadores culturais, opções tradutórias da equipe do IBGE e opções tradutórias alternativas sugeridas pelo consultor

Marcador cultural	Tradução do IBGE	Tradução alternativa
açudes	dams	reservoirs
aglomerado/aglomerados	agglomerate	conurbation
baixadas	slums in glens low-lands baixadas (slums in low-lands) slums in low- lands	low-lands
candomblé	candomblé	candomblé
carteira (assinada/de trabalho assinada)	formal contract employment record card formal employment contract formal labor contract	fully-employed
distritos	districts counties	tradução considerada adequada
favelas	favelas favelas (slums) slums	shanty towns
grotas	caves grotas (slums in deep valleys) slums in deep valleys	caves
igarapés	creeks omissão	tradução considerada adequada
indígena	indian indians indigenous indian-related omissão	indigenous
indígenas	indians indian indigenous omissão	indigenous
logradouro	address places street omissão	-----
logradouros	omissão address areas	-----
maloca	malocas maloca	-----

malocas	malocas	-----
mocambos	mocambo (type of shack) type of shack	tradução considerada adequada
município	municipality municipalities municipal omissão	province
municípios	municipalities municipal sem tradução* municipality omissão	provinces
pardo	brown	tradução considerada adequada
pardos	brown browns	tradução considerada adequada
região	region omissão major regions regions	tradução considerada adequada
regiões	regions major regions omissão do marcador areas	tradução considerada adequada
grandes regiões	major regions regions areas	-----
região metropolitana	metropolitan area metropolitan areas	tradução considerada adequada
regiões metropolitanas	metropolitan area metropolitan areas	tradução considerada adequada
ressacas	slums in back waters ressacas (slums in back waters)	slums in backwaters
umbanda	umbanda	-----

APÊNDICE D – Data de publicação e títulos dos *releases* selecionados para o corpus de estudo

Data de Publicação	Título em português	Título em inglês
05/05/2010	IBGE inicia contagem regressiva para o Censo 2010	<i>IBGE starts countdown for the 2010 Census</i>
19/07/2010	Censo 2010: recenseadores entram em etapa final de treinamento	<i>Census 2010: enumerators start final training phase</i>
26/07/2010	IBGE realiza seminário sobre Censo 2010 para jornalistas	<i>IBGE hosts a seminar about the 2010 Census for journalists</i>
30/07/2010	Navegando por rios ou via satélite: como as entrevistas do Censo 2010 chegam ao IBGE	<i>River and satellite navigation put 2010 Census interviews on the way to IBGE</i>
02/08/2010	Começa o Censo Demográfico 2010	<i>2010 Demographic Census begins</i>
16/08/2010	Censo 2010: 17% da população já foram recenseados	<i>2010 Census: 17% of the population has been enumerated</i>
18/08/2010	Censo 2010: IBGE faz seleção complementar para mais 2.006 recenseadores	<i>2010 Census: IBGE selects 2,006 more enumerators</i>
30/08/2010	Censo 2010: 48% da população já foram recenseados	<i>2010 Census: 48% of the population has been counted</i>
27/09/2010	Dados preliminares do Censo 2010 já revelam	<i>Preliminary data of the 2010 Census already</i>

	mudanças na pirâmide etária brasileira	<i>reveal changes in the Brazilian population pyramid</i>
29/09/2010	Seminário sobre metodologia e tecnologia do Censo 2010 reúne representantes de 20 países	<i>Seminar on the 2010 Census methodology and technology brings together representatives of 20 countries</i>
04/11/2010	IBGE divulga os resultados da coleta do Censo 2010	<i>IBGE releases results of the 2010 Census collection</i>
29/11/2010	Censo 2010: população do Brasil é de 190.732.694 pessoas	<i>2010 Census: Brazilian population amounts to 190,732,694 persons</i>
28/01/2011	Unesco premia IBGE por tecnologia no Censo 2010	<i>UNESCO rewards IBGE for technology in the 2010 Census</i>
29/04/2011	Primeiros resultados definitivos do Censo 2010: população do Brasil é de 190.755.799 pessoas	<i>2010 Census first final results: Brazil has a population of 190,755,799 residents</i>
01/07/2011	IBGE divulga Malha Municipal e Informações dos Setores Censitários do Censo 2010	<i>2010 Census: IBGE releases Municipal Grid and Information on Enumeration Areas</i>
16/11/2011	Censo 2010: Mais da metade dos emigrantes brasileiros são mulheres	<i>2010 Census: More than half of Brazilian emigrants are women</i>
16/11/2011	Censo 2010: País tem declínio de fecundidade e migração e aumentos na	<i>2010 Census: Country faces decline of fertility and sees advances in schooling, employment</i>

	escolarização, ocupação e posse de bens duráveis	<i>and ownership of durable goods</i>
16/11/2011	Indicadores Sociais Municipais 2010: incidência de pobreza é maior nos municípios de porte médio	<i>2010 Municipal Social Indicators : incidence of poverty is higher in medium size municipalities</i>
25/11/2011	Cadastro Nacional de Endereços para Fins Estatísticos auxiliará na produção de pesquisas domiciliares	<i>National Address File for Statistical Purposes will support household surveys</i>
15/12/2011	Censo 2010 aprimorou a identificação dos aglomerados subnormais	<i>2010 Census improved identification of subnormal agglomerates</i>
21/12/2011	Censo 2010: 11,4 milhões de brasileiros (6,0%) vivem em aglomerados subnormais	<i>2010 Census: 11.4 million Brazilians (6.0%) live in subnormal agglomerates</i>
27/04/2012	Censo 2010: escolaridade e rendimento aumentam e cai mortalidade infantil	<i>2010 Census: schooling and income increase and infant mortality falls</i>
25/05/2012	Censo 2010 revela: mais da metade dos domicílios situavam-se em locais sem bueiros	<i>2010 Census reveals: more than half of the households were placed in localities without manholes/storm drains</i>
12/06/2012	Censo 2010: Mais de 3 milhões de crianças e adolescentes trabalhavam no Brasil	<i>2010 Census: Brazil has more than 3 million working children and teenagers</i>

29/06/2012	Censo 2010: número de católicos cai e aumenta o de evangélicos, espíritas e sem religião	<i>2010 Census: number of Catholics falls and number of Protestants, Spiritists and persons without religion records increase</i>
10/08/2012	Censo 2010: população indígena é de 896,9 mil, tem 305 etnias e fala 274 idiomas	<i>2010 Census: population of Indians is formed by 896.9 thousand persons, has 305 ethnicities and speaks 274 languages</i>
17/10/2012	Censo 2010: Uniões consensuais já representam mais de 1/3 dos casamentos e são mais frequentes nas classes de menor rendimento	<i>2010 Census: Consensual unions make up more than 1/3 of marriages and are more common among lower classes</i>
19/12/2012	Censo 2010: mulheres são mais instruídas que homens e ampliam nível de ocupação	<i>2010 Census: women have higher schooling than men and expand their level of employment</i>
27/05/2013	Censo 2010: IBGE apresenta mapa de indicadores sobre trabalho infantil	<i>2010 Census: IBGE presents map of child labor indicators in Brazil</i>
28/06/2013	Atlas do Censo Demográfico do IBGE mapeia mudanças na sociedade brasileira	<i>Atlas of the Population Census of IBGE maps changes in the Brazilian society</i>
30/09/2013	Censo 2010: IBGE apresenta indicadores sobre trabalho infantil	<i>2010 Census: IBGE presents child labor indicators</i>

06/11/2013	Censo 2010 mostra as características territoriais dos aglomerados subnormais e suas diferenças das demais áreas das cidades	<i>2010 Census reveals territorial characteristics of subnormal agglomerates and differences from other city areas</i>
16/03/2013	Grade Estatística permite obter dados do Censo 2010 para diversos recortes espaciais	<i>Statistical grid allows applying 2010 Census data to several spatial divisions</i>

APÊNDICE E – Ano de publicação e títulos dos textos do *Statistics Canada*

Ano de publicação	Título do texto
2012	<i>Living arrangements of seniors</i>
2012	<i>The Canadian Population in 2011: Population Counts and Growth</i>
2012	<i>Linguistic Characteristics of Canadians</i>
2013	<i>Income Composition in Canada</i>
2013	<i>Immigration and Ethnocultural Diversity in Canada</i>
2013	<i>Education in Canada: Attainment, Field of Study and Location of Study</i>
2012	<i>Portrait of Families and Living Arrangements in Canada</i>
2012	<i>The Canadian Population in 2011: Age and Sex</i>
2013	<i>Aboriginal Peoples in Canada: First Nations People, Métis and Inuit</i>
2013	<i>Home Ownership and Shelter Costs</i>

APÊNDICE F – Ano de publicação e títulos dos textos do *US Bureau of Statistics*

Ano de publicação	Título do texto
01/01/2010	<i>Census Bureau Signs 150,000th 2010 Census Partner</i>
04/01/2010	<i>Census Bureau Launches 2010 Census Road Tour Across Nation</i>
07/02/2010	<i>2010 Census Ad Campaign Positioned to Save Taxpayers Millions Helping Boost Mail Back Participation Rates</i>
13/01/2010	<i>Census Bureau, Lockheed Martin Open Utah Call Center</i>
14/01/2010	<i>Census Bureau Launches 2010 Census Advertising Campaign</i>
25/01/2010	<i>2010 Census Gets Under Way in Remote Alaska</i>
12/02/2010	<i>2010 Census Ads Set to Air During Winter Olympics</i>
23/02/2010	<i>National and Local Organizations Help Increase Awareness as 2010 Census Partnerships Grow to 200,000</i>
25/02/2010	<i>Census Bureau Launches Online Mapping Tool Showing 2000 Census Participation Rates to Help Communities Prepare for 2010 Census</i>
01/03/2010	<i>Census Takers Begin Hand Delivering 2010 Census Questionnaires to 12 Million U.S. Addresses</i>
02/03/2010	<i>Census Bureau Makes Special Efforts for Complete Count of Hurricane Affected Areas in the Gulf Coast</i>

08/03/2010	<i>120 Million Households to Begin Receiving 2010 Census Advance Letter</i>
09/03/2010	<i>Census Bureau Launches "Children Count Too" Awareness Campaign Featuring Nickelodeon's Dora the Explorer</i>
15/03/2010	<i>2010 Census Forms Arrive in 120 Million Mailboxes Across Nation</i>
16/03/2010	<i>Census Bureau Back at the Track for Second Race with Biffle's No. 16 Ford Fusion</i>
16/03/2010	<i>Census Bureau Director Delivers 2010 Census Message from James Madison's Montpelier Home</i>
18/03/2010	<i>Census Bureau Unveils 20-Foot 2010 Census Forms in Major Cities</i>
24/03/2010	<i>Census Bureau Rolls Out New Interactive Maps Showing 2010 Census Mail Participation Rates</i>
24/03/2010	<i>U.S. Census Bureau Delivers Final State 2010 Census Population Totals for Legislative Redistricting</i>
25/03/2010	<i>Census Bureau Wraps Up Three-Race Roush Fenway Sponsorship With a Trip to the 'Clip</i>
06/04/2010	<i>Second Round of Census Forms Mailed to 40 Million Households</i>
08/04/2010	<i>Brewers Pitcher Gallardo and Other Major League Players Remind Residents It's Not Too Late to Mail Back Census Forms</i>
15/04/2010	<i>Two Days Left to Mail Back 2010 Census Forms</i>

23/04/2010	<i>America Matches Mail Participation Rate from 2000 Census</i>
28/04/2010	<i>The Numbers Are In: 72 Percent of Nation's Households Mail Back 2010 Census Forms</i>
30/04/2010	<i>Door-to-Door Visits Begin for 2010 Census</i>
04/05/2010	<i>King Family Urges Residents to Open Their Doors and Answer the 2010 Census</i>
07/05/2010	<i>2010 Census Advertising Named Multicultural and Best Branded Campaign of the Year</i>
08/06/2010	<i>Census Bureau Takes Steps to Improve Accuracy of 2010 Census in Mississippi Delta</i>
07/07/2010	<i>Census Bureau Director Reports 2010 Census on Schedule, Under Budget</i>
30/07/2010	<i>Last Day to Ensure You Are Included in 2010 Census</i>
10/08/2010	<i>\$1.6 Billion in 2010 Census Savings Returned</i>
06/10/2010	<i>2010 Census Shows Second Highest Homeownership Rate on Record Despite Largest Decrease since 1940</i>
19/10/2010	<i>Census Bureau Associate Director Elected to Prestigious Institute of Medicine</i>
21/12/2010	<i>U.S. Census Bureau Announces 2010 Census Population Counts Apportionment Counts Delivered to President</i>

29/01/2011	<i>2010 Census Shows Black Population has Highest Concentration in the South</i>
24/03/2011	<i>2010 Census Shows America's Diversity</i>
24/03/2011	<i>U.S. Census Bureau Releases Data on Population Distribution and Change in the U.S. Based on Analysis of 2010 Census Results</i>
31/03/2011	<i>2010 Census Advertising Campaign Earns Prestigious Ogilvy Awards</i>
30/11/2011	<i>2010 Census Shows 65 and Older Population Growing Faster Than Total U.S. Population</i>
25/06/2010	<i>Brooklyn Households May Get Additional Visit From Census Bureau</i>
25/10/2010	<i>Census Bureau Names Behler as Director of Detroit Regional Office</i>
20/04/2011	<i>2010 Census Advance Group Quarters Summary File</i>
09/05/2011	<i>Plato, Mo. Celebrates Recognition as the 2010 Census U.S. Center of Population</i>
11/05/2011	<i>Census Bureau Wins Seven Awards from the National Association of Government Communicators</i>
26/05/2011	<i>2010 Census Shows Nation's Hispanic Population Grew Four Times Faster Than Total U.S. Population</i>
21/06/2011	<i>Census Bureau Recognized by Computerworld For Customer Service Innovation</i>
26/06/2011	<i>2010 Census Shows Nation's Population is Aging</i>
22/06/2011	<i>2010 Census Campaign Wins Bronze Effie Award</i>

24/08/2011	<i>U.S. Census Bureau Releases 2010 Census Population Counts for the U.S. Virgin Islands</i>
25/08/2011	<i>Census Bureau Releases New Local-Level Demographic Information from 2010 Census for Mississippi, New Mexico, North Dakota, Rhode Island and Tennessee</i>
27/09/2011	<i>Census Bureau Releases Estimates of Same-Sex Married Couples</i>
29/09/2011	<i>2010 Census Shows White Population Growth Fueled by Hispanics</i>
29/09/2011	<i>2010 Census Shows Black Population has Highest Concentration in the South</i>
12/10/2011	<i>Census Bureau Releases 2011 Determinations of Political Jurisdictions Subject to Minority Language Assistance</i>
25/01/2012	<i>2010 Census Shows Nearly Half of American Indians and Alaska Natives Report Multiple Races</i>
21/03/2012	<i>2010 Census Shows Asians are Fastest-Growing Race Group</i>
26/03/2012	<i>Growth in Urban Population Outpaces Rest of Nation, Census Bureau Reports</i>
08/05/2012	<i>2010 Census Shows More than Half of Native Hawaiians and Other Pacific Islanders Report Multiple Races</i>
25/04/2012	<i>2010 Census Shows Interracial and Interethnic Married Couples Grew by 28 Percent over Decade</i>
22/05/2012	<i>Census Bureau Releases Estimates of Undercount and Overcount in the 2010 Census</i>

27/09/2012	<i>Populations Increasing in Many Downtowns, Census Bureau Reports</i>
27/09/2012	<i>2010 Census Shows Multiple-Race Population Grew Faster Than Single-Race Population</i>
27/09/2012	<i>Census Bureau Releases Report on 2010 Census Emergency and Transitional Shelter Population</i>
10/12/2012	<i>2010 Census Report Shows More Than 80 Percent of Centenarians are Women</i>